

OM of: US-10-040-803-7 to: GenEmbl:\* out-format: pfs  
Date: Aug 13, 2002 1:35 PM

About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODEL=frme-p2n.html -DEV=xml  
-Q=/cna2/\_USFTO.spool/US10040803/runat\_13082002.083352.7339/app-query.fasta\_1.378  
-DB=GenEmbl -QFMT=fastap -SUFFIX=p2n.rge -GAPOP=12.000  
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPEL=0.000 -LOOPEXT=0.000  
-OGAPOP=4.500 -OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500  
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500  
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blonsum62  
-TRANS=human40.csi -LIST=45 -DOCALLIGN=200 -THR\_SCORE=pct  
-THR\_MAX=100 -THR\_MIN=0 -ALIGN=10 -MODE=LOCAL -OUTFMT=pfs  
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-R=US10040803.@CGN1\_1.4141 -NCPU=6 -ICPU=3 -LONGLOG  
-TIMEOUT=120 -WARN\_TIMEOUT=30 -NO\_XLPPY -WAIT -THREADS=1

Information block:

Query: US-10-040-803-7  
Query length: 317  
Database: GenEmbl:\*  
Database sequences: 1797656  
Database length: 1873333701  
Search time (sec): 1864.040000

Score list:

Sequence	Strd Orig	Zscore	EScore	Len	Documentation
gb.pr:AB010779	+ 1720.00	2372.81	8.4e-124	1306	AB010779 Homo sapiens hbsssp-4
gb.pr:AB021182	+ 1720.00	2372.64	8.5e-124	1332	AB021182 Homo sapiens, serine P
gb.pr:BC009726	+ 1720.00	2372.20	9.0e-124	1403	BC009726 Homo sapiens, proteas
gb.pr:AB010778	+ 1275.50	1757.56	1.6e-89	1321	AB010778 Mus musculus, hbsssp-4
gb.pr:R05642	+ 1293.00	1685.32	1.6e-85	1259	AJ005642 Rattus rattus mRNA en
gb.pr:AC003965	+ 1094.50	1479.06	5.1e-74	3462	AC003965 Homo sapiens chromos
gb.pr:HS030593	+ 739.00	1017.36	5.5e-42	1013	AJ306593 Homo sapiens mRNA for
gb.ov:AB018694	+ 661.50	903.97	6.8e-38	1796	AB018694 Xenopus laevis Xespin
gb.pr:AB0098193	+ 607.50	830.47	6.8e-38	1809	BC001462 Homo sapiens, proteas
gb.pr:BC001462	+ 607.50	830.29	6.9e-38	1834	AX335777 Sequence 6286 from pa
gb.pr:AX335777	+ 607.50	830.29	6.9e-38	1834	AX335777 Sequence 6286 from pa
gb.pr:AX336076	+ 607.50	830.29	6.9e-38	1834	AX336076 Sequence 6585 from pa
gb.pr:AX336340	+ 607.50	830.29	6.9e-38	1834	AX336340 Sequence 6849 from pa
gb.pr:AX335775	+ 607.50	830.29	6.9e-38	1834	AX335775 Sequence 6286 from pa
gb.pr:AX098215	+ 607.50	830.29	6.9e-38	1835	AX098215 Sequence 127 from pat
gb.pr:AB038496	+ 603.00	832.98	1.6e-37	1850	AB038496 Xenopus laevis Xesp-1
gb.pr:AB02076	+ 589.00	804.55	1.9e-36	1864	AB02076 Rattus norvegicus pnc
gb.pr:AB017638	+ 589.00	803.09	2.3e-36	2208	AB017638 Rattus norvegicus mkn
gb.pr:AB017638	+ 579.50	797.94	4.4e-36	869	Y18224 Ovis aries mRNA for tytr
gb.com:OAR18224	+ 579.50	797.94	4.4e-36	869	Y18224 Ovis aries mRNA for tytr
gb.pr:BC003851	+ 578.50	790.55	1.1e-35	1743	BC003851 Mus musculus, Simlad
gb.pr:AF378085	+ 578.50	790.55	1.1e-35	1753	AF378085 Mus musculus prosta
gb.pr:AB038244	+ 578.50	790.05	1.2e-35	1856	AB038244 Mus musculus mRNA fo
gb.pr:AX342626	+ 575.50	785.11	2.3e-35	2036	AX342626 Sequence 23 from Patc
gb.pr:AF188613	+ 574.50	780.04	2.4e-35	1797	AF188613 Mus musculus channel
gb.pr:AX338508	+ 573.50	784.79	4.4e-35	2662	AX338508 Sequence 27 from Patc
gb.pr:AF175522	+ 571.00	783.42	2.8e-35	1199	AF175522 Homo sapiens transmem
gb.pr:BC010970	+ 571.00	781.62	3.6e-35	1478	BC010970 Mus musculus, Simlad
gb.pr:BC010970	+ 564.00	772.21	1.2e-34	1431	AJ243866 Mus musculus mRNA fo
gb.pr:MM0234366	+ 560.50	770.43	1.5e-34	1002	Y18223 Ovis aries mRNA for tytr
gb.com:OAR18223	+ 560.50	770.43	1.5e-34	1002	Y18223 Ovis aries mRNA for tytr
gb.pr:AX07905	+ 554.50	752.43	1.5e-33	3104	AX07905 Sequence 9 from Paten
gb.pr:AX07903	+ 554.50	752.43	1.5e-33	3104	AX07903 Sequence 7 from Paten
gb.pr:AX07903	+ 554.00	760.71	5.2e-34	1090	AX13853 Mus musculus serine pnc
gb.pr:AMUSM06	+ 554.00	760.57	5.3e-34	1108	AM080464 Sequence 20 from patc
gb.pr:AB080464	+ 554.00	760.57	5.3e-34	1108	AM080464 Sequence 20 from patc
gb.pr:AMUSM06	+ 554.00	760.57	5.3e-34	1108	M57626 Mouse protease-6 gene
gb.pr:AMUSM06	+ 553.00	754.30	1.2e-33	1958	AX132632 Sequence 29 from Patc
gb.pr:AX132632	+ 553.00	754.30	1.2e-33	1958	AX132632 Sequence 29 from Patc
gb.pr:AX080456	+ 546.00	750.42	2.0e-33	995	M24664 Dog mast cell tryptase m
gb.com:DOGC8576	+ 538.50	739.20	8.2e-33	1097	AB080464 Sequence 4 from paten
gb.pr:AB080456	+ 538.50	739.20	8.2e-33	1097	U36755 Rat mRNA for mast cell
gb.pr:R0567909	+ 536.50	736.39	1.2e-32	1103	D57809 Rattus norvegicus mast
gb.com:BT0967909	+ 536.00	737.17	1.1e-32	929	X94982 B.taurus mRNA for tryptase

```

34  rglleProValProProAlaCysGlyLysProGlnGlnLeuAsnArgVal 50
104  GGATACCTGTCCCGACCGCTGTGGAGAGCCCGACAGCTAACCGGGTT 153
51  ValGlyGlyLysSerThrAspSerGluTrpProTrpIleValSerI 67
154  GTGGGGGGGAGGAGACGACGACGAGGAGGAGGAGGAGGAGGAGGAG 203
67  eGlnLysAsnGlyThrHisHisCysAlaGlySerLeuLeuThrSerArg 84
204  CCAGAGAGATGGAGACCCACGCTCCGAGGAGGAGGAGGAGGAGGAGG 253
84  rPValIleThrAlaAlaHisCysPheLysAspAsnLeuAsnLysProTyr 100
254  GGGTGAATCAGTGTGCCACGCTGTTCAGAGCAGACGAGCAACCAACAT 303
101  LeuPheSerValLeuLeuGlyAlaTrpGlnLeuGlnLysProGlySerAr 117
304  CTGTCTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 353
117  gSerGlnLysValGlyAlaAlaTrpValGluProHisProValTyrSer 134
354  GTCCAGAGAGTGGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 403
134  rPlyGlyGlyAlaCysAlaAspIleAlaLeuValArgLeuGluArgSer 150
404  GGAAGGAGAGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 453
151  IleGlnPheSerGluArgValLeuProIleCysLeuProAspAlaSerI 167
454  ATNAGATTCAGAGCGGGTCTGTGTGTGTGTGTGTGTGTGTGTGTGT 503
167  eHisLeuProProAsnThrHisCysTrpIleSerGlyTrpGlySerIle 184
504  CCACCTCTCTCAAAACACCCAGCTGTGTGTGTGTGTGTGTGTGTGTGT 553
184  LAspGlyValProLeuProHisProGlnThrLeuGlnLysLeuVal 200
554  AAGTGTGAGTCTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 603
201  ProIleIleAspSerGluValLysSerHisLeuTrpTrpArgIleVal 217
604  CCAATCATGACCTCGGAGAGTCTGTGTGTGTGTGTGTGTGTGTGTGT 653
217  YGlnGlyProIleThrGluAspMetLeuCysAlaGlyTrpLeuGluGly 234
654  ACAGGGAGCCCATCATGAGGACATGCTGTGTGTGTGTGTGTGTGTGTGT 703
34  LuArgAspAlaCysLeuGlyAspSerGlyLysProLeuMetCysGlnVal 250
704  AGCGGAGTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 753
251  AspGlyAlaTrpLeuLeuAlaGlyIleIleSerTrpGlyGlyLysAl 267
754  GAGCGCGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 803
267  agluArgAsnArgProGlyValTyrIleSerLeuSerAlaHisArgSer 284
804  CGAGCGCAACAGGCGCGGGGTCTACATGACCTGTGTGTGTGTGTGTGT 853
284  rPValGluLysIleValGlnGlyValGlnLeuArgGlyValArgIleGly 300
854  GGGTGAAGAAAGATCGTGCAGAGGGGTGTGTGTGTGTGTGTGTGTGTGT 903
301  GlyGlyAlaLeuArgAlaProSerGlnGlySerGlyAlaAlaAlaArgSe 317
904  GGTGGGGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 953
317  r 317
954  C 954

```

```

seq_name: gb_pr:AF321182
seq_documentation_block:
LOCUS       AF321182                1332 bp    mRNA    linear    PRI 26-DEC-2001
DEFINITION Homo sapiens serine protease PRSS22 mRNA, complete cds.
ACCESSION   AF321182
VERSION     AF321182.1
KEYWORDS    AF321182.1 GI:11386012
SOURCE      human.
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE   1 (bases 1 to 1332)
AUTHORS     Wong, G.W., Yasuda, S., Madhusudhan, M.S., Li, L., Yang, Y.,
            Krill, S.A., Sali, A., and Stevens, R.L.
            Human Trypsin epsilon (PRSS22), a New Member of the Chromosome
            16p13.3 Family of Human Serine Proteases Expressed in Airway
            Epithelial Cells
JOURNAL     J. Biol. Chem. 276 (52), 49169-49182 (2001)
PUBMED      11602603
REFERENCE   2 (bases 1 to 1332)
AUTHORS     Wong, G.W.
TITLE       Direct Submission
SUBMITTER   Submitted (14-NOV-2000) Rheumatology, Immunology and Allergy,
            Brigham and Women's Hospital, Harvard Medical School, 1 Jimmy Fund
            Way, Boston, MA 02115, USA
FEATURES    location/Qualifiers
             1..1332
             /organism="Homo sapiens"
             /db_xref="taxon:9606"
             /chromosome="16"
             /map="16p13.3"
             /tissue-type="pancreas"
             18..971
             /codon_start=1
             /product="serine protease PRSS22"
             /protein_id="AAG35070.1"
             /db_xref="GI:11386013"
             /translation="MYSGAPPALGGGCGCTPTSLILASTAIIAARIPVPPACGK
             OOLNRYGSDSDSDSEWPVIVISQKNGTHHCGSLITSRWVITTAHCFRDNINKPYLF
             SYLGWMOJGNRESRSQKGVAMVEHPVYSKEACADIALVRLERSIQSERVLPF
             CLPDASIHLPNTHWISGWSIQDVPVPHQPTLOKLPILIDSEVCSHLYWRGAGQ
             GPTEDMLCAGYLEGRDCLSDSGGGLMCQYDGMALLAGLISMEGCAERNRPVYI
             SISAHRSWEKTIYGVOLRGRAQGGALAPSPQSGAARS"
BASE COUNT  223 a 445 c 403 g 261 t
ORIGIN
Alignment_scores:
Quality: 1720.00      Length: 317
Ratio: 5.426          Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-10-040-803-7 x AF321182 ..
Align seg 1/1 to: AF321182 from: 1 to: 1332
1  MetValValSerGlyAlaProProAlaLeuGlyGlyLysCysLeuGlyTh 17
18  ATGTGTGTTTGTGAGCGCCCCAGCCCTGTGGGGGGGTGTGTGTGTGTGT 67
17  rPheThrSerLeuLeuLeuAlaSerThrAlaIleLeuAsnAlaAla 34
68  CTTCACCTCCCTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 117
34  rglleProValProProAlaCysGlyLysProGlnGlnLeuAsnArgVal 50
118  GGATACCTGTTCGCCAGCGCTGTGGAGAGCCCGACGAGCTGAACCGGGTT 167
51  ValGlyGlyLysSerThrAspSerGluTrpProTrpIleValSerI 67
168  GTGGGGGCGAGGAGACGACGACGAGGAGGAGGAGGAGGAGGAGGAGGAT 217

```

```

67  eGlnLysasnGlyThrHisCysAlaGlySerLeuLeuThrSerArg 84
68  |||||||
69  |||||||
70  |||||||
71  |||||||
72  |||||||
73  |||||||
74  |||||||
75  |||||||
76  |||||||
77  |||||||
78  |||||||
79  |||||||
80  |||||||
81  |||||||
82  |||||||
83  |||||||
84  rPValIleThrAlaAlaHisCysPheLysAspAsnLeuAsnLysProTyr 100
85  |||||||
86  |||||||
87  |||||||
88  |||||||
89  |||||||
90  |||||||
91  |||||||
92  |||||||
93  |||||||
94  |||||||
95  |||||||
96  |||||||
97  |||||||
98  |||||||
99  |||||||
100  |||||||
101  LeuPheSerValLeuLeuGlyAlaTrpGlnLeuGlyLysnProGlySerAr 117
102  |||||||
103  |||||||
104  |||||||
105  |||||||
106  |||||||
107  |||||||
108  |||||||
109  |||||||
110  |||||||
111  |||||||
112  |||||||
113  |||||||
114  |||||||
115  |||||||
116  |||||||
117  gSerGlnLysValGlyAlaTrpValAlaTrpProHisProValTyrSer 134
118  |||||||
119  |||||||
120  |||||||
121  |||||||
122  |||||||
123  |||||||
124  |||||||
125  |||||||
126  |||||||
127  |||||||
128  |||||||
129  |||||||
130  |||||||
131  |||||||
132  |||||||
133  |||||||
134  rPlySGlnGlyAlaCysAlaAspIleAlaLeuValArgLeuGlyArgSer 150
135  |||||||
136  |||||||
137  |||||||
138  |||||||
139  |||||||
140  |||||||
141  |||||||
142  |||||||
143  |||||||
144  |||||||
145  |||||||
146  |||||||
147  |||||||
148  |||||||
149  |||||||
150  |||||||
151  |||||||
152  |||||||
153  |||||||
154  |||||||
155  |||||||
156  |||||||
157  |||||||
158  |||||||
159  |||||||
160  |||||||
161  |||||||
162  |||||||
163  |||||||
164  |||||||
165  |||||||
166  |||||||
167  |||||||
168  |||||||
169  |||||||
170  |||||||
171  |||||||
172  |||||||
173  |||||||
174  |||||||
175  |||||||
176  |||||||
177  |||||||
178  |||||||
179  |||||||
180  |||||||
181  |||||||
182  |||||||
183  |||||||
184  |||||||
185  |||||||
186  |||||||
187  |||||||
188  |||||||
189  |||||||
190  |||||||
191  |||||||
192  |||||||
193  |||||||
194  |||||||
195  |||||||
196  |||||||
197  |||||||
198  |||||||
199  |||||||
200  |||||||
201  |||||||
202  |||||||
203  |||||||
204  |||||||
205  |||||||
206  |||||||
207  |||||||
208  |||||||
209  |||||||
210  |||||||
211  |||||||
212  |||||||
213  |||||||
214  |||||||
215  |||||||
216  |||||||
217  |||||||
218  |||||||
219  |||||||
220  |||||||
221  |||||||
222  |||||||
223  |||||||
224  |||||||
225  |||||||
226  |||||||
227  |||||||
228  |||||||
229  |||||||
230  |||||||
231  |||||||
232  |||||||
233  |||||||
234  |||||||
235  |||||||
236  |||||||
237  |||||||
238  |||||||
239  |||||||
240  |||||||
241  |||||||
242  |||||||
243  |||||||
244  |||||||
245  |||||||
246  |||||||
247  |||||||
248  |||||||
249  |||||||
250  |||||||
251  |||||||
252  |||||||
253  |||||||
254  |||||||
255  |||||||
256  |||||||
257  |||||||
258  |||||||
259  |||||||
260  |||||||
261  |||||||
262  |||||||
263  |||||||
264  |||||||
265  |||||||
266  |||||||
267  |||||||
268  |||||||
269  |||||||
270  |||||||
271  |||||||
272  |||||||
273  |||||||
274  |||||||
275  |||||||
276  |||||||
277  |||||||
278  |||||||
279  |||||||
280  |||||||
281  |||||||
282  |||||||
283  |||||||
284  |||||||
285  |||||||
286  |||||||
287  |||||||
288  |||||||
289  |||||||
290  |||||||
291  |||||||
292  |||||||
293  |||||||
294  |||||||
295  |||||||
296  |||||||
297  |||||||
298  |||||||
299  |||||||
300  |||||||
301  |||||||
302  |||||||
303  |||||||
304  |||||||
305  |||||||
306  |||||||
307  |||||||
308  |||||||
309  |||||||
310  |||||||
311  |||||||
312  |||||||
313  |||||||
314  |||||||
315  |||||||
316  |||||||
317  |||||||
318  |||||||
319  |||||||
320  |||||||
321  |||||||
322  |||||||
323  |||||||
324  |||||||
325  |||||||
326  |||||||
327  |||||||
328  |||||||
329  |||||||
330  |||||||
331  |||||||
332  |||||||
333  |||||||
334  |||||||
335  |||||||
336  |||||||
337  |||||||
338  |||||||
339  |||||||
340  |||||||
341  |||||||
342  |||||||
343  |||||||
344  |||||||
345  |||||||
346  |||||||
347  |||||||
348  |||||||
349  |||||||
350  |||||||
351  |||||||
352  |||||||
353  |||||||
354  |||||||
355  |||||||
356  |||||||
357  |||||||
358  |||||||
359  |||||||
360  |||||||
361  |||||||
362  |||||||
363  |||||||
364  |||||||
365  |||||||
366  |||||||
367  |||||||
368  |||||||
369  |||||||
370  |||||||
371  |||||||
372  |||||||
373  |||||||
374  |||||||
375  |||||||
376  |||||||
377  |||||||
378  |||||||
379  |||||||
380  |||||||
381  |||||||
382  |||||||
383  |||||||
384  |||||||
385  |||||||
386  |||||||
387  |||||||
388  |||||||
389  |||||||
390  |||||||
391  |||||||
392  |||||||
393  |||||||
394  |||||||
395  |||||||
396  |||||||
397  |||||||
398  |||||||
399  |||||||
400  |||||||
401  |||||||
402  |||||||
403  |||||||
404  |||||||
405  |||||||
406  |||||||
407  |||||||
408  |||||||
409  |||||||
410  |||||||
411  |||||||
412  |||||||
413  |||||||
414  |||||||
415  |||||||
416  |||||||
417  |||||||
418  |||||||
419  |||||||
420  |||||||
421  |||||||
422  |||||||
423  |||||||
424  |||||||
425  |||||||
426  |||||||
427  |||||||
428  |||||||
429  |||||||
430  |||||||
431  |||||||
432  |||||||
433  |||||||
434  |||||||
435  |||||||
436  |||||||
437  |||||||
438  |||||||
439  |||||||
440  |||||||
441  |||||||
442  |||||||
443  |||||||
444  |||||||
445  |||||||
446  |||||||
447  |||||||
448  |||||||
449  |||||||
450  |||||||
451  |||||||
452  |||||||
453  |||||||
454  |||||||
455  |||||||
456  |||||||
457  |||||||
458  |||||||
459  |||||||
460  |||||||
461  |||||||
462  |||||||
463  |||||||
464  |||||||
465  |||||||
466  |||||||
467  |||||||
468  |||||||
469  |||||||
470  |||||||
471  |||||||
472  |||||||
473  |||||||
474  |||||||
475  |||||||
476  |||||||
477  |||||||
478  |||||||
479  |||||||
480  |||||||
481  |||||||
482  |||||||
483  |||||||
484  |||||||
485  |||||||
486  |||||||
487  |||||||
488  |||||||
489  |||||||
490  |||||||
491  |||||||
492  |||||||
493  |||||||
494  |||||||
495  |||||||
496  |||||||
497  |||||||
498  |||||||
499  |||||||
500  |||||||
501  |||||||
502  |||||||
503  |||||||
504  |||||||
505  |||||||
506  |||||||
507  |||||||
508  |||||||
509  |||||||
510  |||||||
511  |||||||
512  |||||||
513  |||||||
514  |||||||
515  |||||||
516  |||||||
517  |||||||
518  |||||||
519  |||||||
520  |||||||
521  |||||||
522  |||||||
523  |||||||
524  |||||||
525  |||||||
526  |||||||
527  |||||||
528  |||||||
529  |||||||
530  |||||||
531  |||||||
532  |||||||
533  |||||||
534  |||||||
535  |||||||
536  |||||||
537  |||||||
538  |||||||
539  |||||||
540  |||||||
541  |||||||
542  |||||||
543  |||||||
544  |||||||
545  |||||||
546  |||||||
547  |||||||
548  |||||||
549  |||||||
550  |||||||
551  |||||||
552  |||||||
553  |||||||
554  |||||||
555  |||||||
556  |||||||
557  |||||||
558  |||||||
559  |||||||
560  |||||||
561  |||||||
562  |||||||
563  |||||||
564  |||||||
565  |||||||
566  |||||||
567  |||||||
568  |||||||
569  |||||||
570  |||||||
571  |||||||
572  |||||||
573  |||||||
574  |||||||
575  |||||||
576  |||||||
577  |||||||
578  |||||||
579  |||||||
580  |||||||
581  |||||||
582  |||||||
583  |||||||
584  |||||||
585  |||||||
586  |||||||
587  |||||||
588  |||||||
589  |||||||
590  |||||||
591  |||||||
592  |||||||
593  |||||||
594  |||||||
595  |||||||
596  |||||||
597  |||||||
598  |||||||
599  |||||||
600  |||||||
601  |||||||
602  |||||||
603  |||||||
604  |||||||
605  |||||||
606  |||||||
607  |||||||
608  |||||||
609  |||||||
610  |||||||
611  |||||||
612  |||||||
613  |||||||
614  |||||||
615  |||||||
616  |||||||
617  |||||||
618  |||||||
619  |||||||
620  |||||||
621  |||||||
622  |||||||
623  |||||||
624  |||||||
625  |||||||
626  |||||||
627  |||||||
628  |||||||
629  |||||||
630  |||||||
631  |||||||
632  |||||||
633  |||||||
634  |||||||
635  |||||||
636  |||||||
637  |||||||
638  |||||||
639  |||||||
640  |||||||
641  |||||||
642  |||||||
643  |||||||
644  |||||||
645  |||||||
646  |||||||
647  |||||||
648  |||||||
649  |||||||
650  |||||||
651  |||||||
652  |||||||
653  |||||||
654  |||||||
655  |||||||
656  |||||||
657  |||||||
658  |||||||
659  |||||||
660  |||||||
661  |||||||
662  |||||||
663  |||||||
664  |||||||
665  |||||||
666  |||||||
667  |||||||
668  |||||||
669  |||||||
670  |||||||
671  |||||||
672  |||||||
673  |||||||
674  |||||||
675  |||||||
676  |||||||
677  |||||||
678  |||||||
679  |||||||
680  |||||||
681  |||||||
682  |||||||
683  |||||||
684  |||||||
685  |||||||
686  |||||||
687  |||||||
688  |||||||
689  |||||||
690  |||||||
691  |||||||
692  |||||||
693  |||||||
694  |||||||
695  |||||||
696  |||||||
697  |||||||
698  |||||||
699  |||||||
700  |||||||
701  |||||||
702  |||||||
703  |||||||
704  |||||||
705  |||||||
706  |||||||
707  |||||||
708  |||||||
709  |||||||
710  |||||||
711  |||||||
712  |||||||
713  |||||||
714  |||||||
715  |||||||
716  |||||||
717  |||||||
718  |||||||
719  |||||||
720  |||||||
721  |||||||
722  |||||||
723  |||||||
724  |||||||
725  |||||||
726  |||||||
727  |||||||
728  |||||||
729  |||||||
730  |||||||
731  |||||||
732  |||||||
733  |||||||
734  |||||||
735  |||||||
736  |||||||
737  |||||||
738  |||||||
739  |||||||
740  |||||||
741  |||||||
742  |||||||
743  |||||||
744  |||||||
745  |||||||
746  |||||||
747  |||||||
748  |||||||
749  |||||||
750  |||||||
751  |||||||
752  |||||||
753  |||||||
754  |||||||
755  |||||||
756  |||||||
757  |||||||
758  |||||||
759  |||||||
760  |||||||
761  |||||||
762  |||||||
763  |||||||
764  |||||||
765  |||||||
766  |||||||
767  |||||||
768  |||||||
769  |||||||
770  |||||||
771  |||||||
772  |||||||
773  |||||||
774  |||||||
775  |||||||
776  |||||||
777  |||||||
778  |||||||
779  |||||||
780  |||||||
781  |||||||
782  |||||||
783  |||||||
784  |||||||
785  |||||||
786  |||||||
787  |||||||
788  |||||||
789  |||||||
790  |||||||
791  |||||||
792  |||||||
793  |||||||
794  |||||||
795  |||||||
796  |||||||
797  |||||||
798  |||||||
799  |||||||
800  |||||||
801  |||||||
802  |||||||
803  |||||||
804  |||||||
805  |||||||
806  |||||||
807  |||||||
808  |||||||
809  |||||||
810  |||||||
811  |||||||
812  |||||||
813  |||||||
814  |||||||
815  |||||||
816  |||||||
817  |||||||
818  |||||||
819  |||||||
820  |||||||
821  |||||||
822  |||||||
823  |||||||
824  |||||||
825  |||||||
826  |||||||
827  |||||||
828  |||||||
829  |||||||
830  |||||||
831  |||||||
832  |||||||
833  |||||||
834  |||||||
835  |||||||
836  |||||||
837  |||||||
838  |||||||
839  |||||||
840  |||||||
841  |||||||
842  |||||||
843  |||||||
844  |||||||
845  |||||||
846  |||||||
847  |||||||
848  |||||||
849  |||||||
850  |||||||
851  |||||||
852  |||||||
853  |||||||
854  |||||||
855  |||||||
856  |||||||
857  |||||||
858  |||||||
859  |||||||
860  |||||||
861  |||||||
862  |||||||
863  |||||||
864  |||||||
865  |||||||
866  |||||||
867  |||||||
868  |||||||
869  |||||||
870  |||||||
871  |||||||
872  |||||||
873  |||||||
874  |||||||
875  |||||||
876  |||||||
877  |||||||
878  |||||||
879  |||||||
880  |||||||
881  |||||||
882  |||||||
883  |||||||
884  |||||||
885  |||||||
886  |||||||
887  |||||||
888  |||||||
889  |||||||
890  |||||||
891  |||||||
892  |||||||
893  |||||||
894  |||||||
895  |||||||
896  |||||||
897  |||||||
898  |||||||
899  |||||||
900  |||||||
901  |||||||
902  |||||||
903  |||||||
904  |||||||
905  |||||||
906  |||||||
907  |||||||
908  |||||||
909  |||||||
910  |||||||
911  |||||||
912  |||||||
913  |||||||
914  |||||||
915  |||||||
916  |||||||
917  |||||||
918  |||||||
919  |||||||
920  |||||||
921  |||||||
922  |||||||
923  |||||||
924  |||||||
925  |||||||
926  |||||||
927  |||||||
928  |||||||
929  |||||||
930  |||||||
931  |||||||
932  |||||||
933  |||||||
934  |||||||
935  |||||||
936  |||||||
937  |||||||
938  |||||||
939  |||||||
940  |||||||
941  |||||||
942  |||||||
943  |||||||
944  |||||||
945  |||||||
946  |||||||
947  |||||||
948  |||||||
949  |||||||
950  |||||||
951  |||||||
952  |||||||
953  |||||||
954  |||||||
955  |||||||
956  |||||||
957  |||||||
958  |||||||
959  |||||||
960  |||||||
961  |||||||
962  |||||||
963  |||||||
964  |||||||
965  |||||||
966  |||||||
967  |||||||
968  |||||||
969  |||||||
970  |||||||
971  |||||||
972  |||||||
973  |||||||
974  |||||||
975  |||||||
976  |||||||
977  |||||||
978  |||||||
979  |||||||
980  |||||||
981  |||||||
982  |||||||
983  |||||||
984  |||||||
985  |||||||
986  |||||||
987  |||||||
988  |||||||
989  |||||||
990  |||||||
991  |||||||
992  |||||||
993  |||||||
994  |||||||
995  |||||||
996  |||||||
997  |||||||
998  |||||||
999  |||||||
1000  |||||||

```

```

seq_name: gb-pr:BC009726
LOCUS      BC009726          1403 bp    mRNA          linear    PRI 22-OCT-2001
DEFINITION Homo sapiens, protease, serine, 22, clone MGC:9599 IMAGE:3899480,
            mRNA, complete cds.
ACCESSION  BC009726
VERSION    BC009726.1 GI:16307274

```

KEYWORDS MGC.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
REFERENCE 1 (bases 1 to 1403)  
AUTHORS Strausberg, R.  
TITLE Direct Submission  
JOURNAL Submitted (29-JUN-2001) National Institutes of Health, Mammalian  
Gene Collection (MGC), Cancer Genomics Office, National Cancer  
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
USA

REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
Contact: MGC help desk  
Email: [cgabs-remail.nih.gov](mailto:cgabs-remail.nih.gov)  
Tissue Procurement: ATCC  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Sequencing Group at the Stanford Human Genome  
Center, Stanford University School of Medicine, Stanford, CA 94305  
Web site: <http://www.shgc.stanford.edu>  
Contact: (Dickson, Mark) [mcdpaxil.stanford.edu](mailto:mcdpaxil.stanford.edu)  
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,  
R. M.

FEATURES  
source  
Location/Qualifiers  
1..1403  
/organism="Homo sapiens"  
/db\_xref="LocusID:64063"  
/db\_xref="taxon:9606"  
/clone="MGC:9599 IMAGE:3899480"  
/tissue\_type="Pancreas, epithelioid carcinoma"  
/clone\_lib="NTH\_MGC\_70"  
/lab\_host="DH10B"  
/note="Vector: pCMV-SPORT6"  
39..992  
/codon\_start=1  
/product="protease, serine, 22"  
/protein\_id="AAH09726.1"  
/db\_xref="GI:16307275"  
/translation="MVSQAPALGGGGLGFTSLILASTAILNARIPIYPAPCKP  
COLNRYVGEDSTDEMPVIVISIQKNGTHHCASLITRWYITAHCKRDMNRPYL  
SVLIGAMOLGNPGRSOKGVAMVEPPEVYSWKEGACADIALVRLERSIQSERVLP  
CLPDASTILPPTWHQISGWSIQDGVLPDLPDOKLXPLIDSEVCSILYMRGAGQ  
GPTIDMICAATLBERDPCAGDSGSLMCOVDAMILLAGIISNGEGCAENRPQVYI  
SISAHRSWVERKIVGVQLRGRAGGGLRALRPSQSGAARAS"

## CDS

BASE COUNT 275 a 452 c 410 g 266 t  
ORIGIN

alignment\_scores:  
Quality: 1720.00 Length: 317  
Ratio: 5.426 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-10-040-803-7 x BC009726 ..

Align seg 1/1 to: BC009726 from: 1 to: 1403

```

1 MetValValSerGlyAlaProProAlaLeuGlyGlyCysLeuGlyThr 17
|||||
39 ATGTTGTTTGTGAGAGCCGCCAGCCCTGGGTGGGGCTGTCTGGCAC 88
|||||
17 rPheThrSerLeuLeuLeuAlaSerThrAlaIleLeuAsnAlaIla 34
|||||
89 CTCACCTCCCTGCTGCTGGCTGGCTGACAGCCATCTCAATGCGGCA 138
|||||

```

```

34  rglleProValProProAlaCysGlyLysProGlnGlnLeuAnaArgVal 50
139  GATATACCTGTTTCCCGACGCTGTGGAGAGCCCGACAGCTGAACCGGGTT 188
51  ValGlyGlyLysAspSerThrAspSerGluTrpProTrrPleValSer1 67
189  GTGGGGGGGAGAGACAGACTGACAGCGAGTGGCCCTGGATGCTGAGCAT 238
67  eGlnLysAsnGlyThrHisHisCysAlaGlySerLeuLeuThrSerArgT 84
239  CAGAGAAAGTGGAGACCCACACCTGCGGAGTTCTGCTGACACAGCGCT 288
84  rPValIleThrAlaAlaHisCysPheLysAspAnLeuAnLysProTyr 100
289  GGTGTATCACTGCTGCCACTGTTTCAAGACCACTGAACCAACCATAC 338
101  LeuPheSerValLeuLeuGlyAlaTrpGlnLeuGlyAspProGlySerX 117
339  CTCTTCTCTGCTGCTGGGGGCTGGAGCTGGGAGACCTGGCTCTCG 388
17  gSerGlnLysValGlyValAlaTrpValGluProHisProValTyrSerT 134
89  GTCCCAAGAGGTGGGTGGTGGTGGTGGAGCCCGACCTGTATCTCT 438
134  rPlySeluGlyAlaCysAlaAspIleAlaLeuValArgLeuGluArgSer 150
439  GAGAGAGAGTGGCTGTGCAGACATGCCCTGGTGGCTGCAGAGCGCTC 488
151  IleGlnPheSerGluArgValLeuProIleCysLeuProAspAlaSer1 167
489  ATACATGTTCTCAGAGCGGTGCTGCCATCTGCTACCTGATGCTCTAT 538
167  eHisLeuProProAsnThrHisCysTrpPheSerGlyTrpGlySer1IleG 184
539  CCACCTCCCTCCCAACACCCACCTGCTGATCTCAGGCTGGGAGAGATCC 588
184  IAspGlyValProLeuProHisProGlnThrLeuGlnLysLeuVal 200
589  AAGATGAGTCCCTTGGCCCAACCTCAGACCTCGAGAGGTGAAGGTT 638
201  ProIleIleAspSerGluValCysSerHisLeuTyrTrpArgGlyValA 217
639  CCATATGATGACCTCGAGAGTCTGCACCCATCTGTACTGGCGGGAGCAG 688
217  yGlnGlyProIleThrGluAspMetLeuCysAlaGlyTyrLeuGluGly 234
689  ACAGGAGACCCATCATGAGACATGCTGTGCGGCTACTTGAGAGGGG 738
34  IuArgAspAlaCysLeuGlyAspSerGlyGlyProLeuMetCysGlnVal 250
139  AGCGGATGCTGTGCTGGGAGCTCGGGGGGCCCTCTCATGTGCGAGGTG 788
251  AspGlyAlaTrpLeuLeuAlaGlyIleIleSerTrpGlyGluGlyCysAl 267
789  GACGGGCGCTGGCTGTGCGCGCATCATCAGCTGGGGGAGAGGCTGTGC 838
267  aGluArgAsnArgProGlyValTyrIleSerLeuSerAlaHisArgSerT 284
839  CGAGGCAACAGGCGCGGGCTGTACATCAGCTCTGCGGACCCCTCT 888
284  rPValGluLysIleValGlnGlyValGlnLeuArgIleArgAlaGlnGly 300
889  GGGTGGAGAGATGCTGCGAAGGGGTGACAGCTCGCGGGCGGCTCAGGGG 938
301  GlyGlyAlaLeuArgAlaProSerGlnGlySerGlyAlaAlaAlaArgSe 317
939  GGTGGGGGCTCAGGAGCAGCAGGAGGCTGTGGGGCGCGCGCGCTC 988
317  r 317
989  C 989
seq_name: gb_AB010778

```

```

seq_documentation_block:
LOCUS      AB010778                1321 bp    mRNA       linear   ROD 17-MAY-2001
DEFINITION Mus musculus MBSP-4 mRNA for brain specific serine protease-4,
complete cds.
ACCESSION  AB010778
VERSION    AB010778.1  GI:12248748
KEYWORDS   Mus musculus brain cDNA to mRNA.
SOURCE     Mus musculus
ORGANISM   Mus musculus
REFERENCE  1 (sites)
AUTHORS   Mitsui,S., Okui,A., Kominami,K. and Yamaguchi,N.
TITLE     Cloning and characterization of a novel serine protease, MBSP-4
JOURNAL   Unpublished
REFERENCE  2 (bases 1 to 1321)
AUTHORS   Yamaguchi,N. and Mitsui,S.
TITLE     Direct Submission
JOURNAL   Submitted (27-JAN-1998) Nozomi Yamaguchi, Institute for Geriatrics,
Kyoto Prefectural University of Medicine, Department of Cell
Biology, Kajii-cho 465, Kamigyo-ku, Kawaramati Hirokoji, Kyoto,
Kyoto 602-8566, Japan (E-mail: nozomi@kyoto.kyu-u.ac.jp,
Tel:81-75-251-5848(ex.5848), Fax:81-75-251-5848)
FEATURES
     source
         1..1321
             /organism="Mus musculus"
             /db_xref="taxon:10090"
             /tissue_type="brain"
     gene
         80..1000
             /gene="MBSP-4"
             /note="prosemin"
     CDS
         80..1000
             /gene="MBSP-4"
             /note="prosemin"
             /codon_start=1
             /product="brain specific serine protease-4"
             /protein_id="BAB20262.1"
             /db_xref="GI:12248749"
             /translation="MMISRPPLAGDGFSLILVLTSTAPISAATIRVSPDGRP
OQINRIVGGEDSDAOWPIVSIILKNGSHHCAGSLITRNVYTAHCFKSNMDSLP
SVLGAMKRGSGSPRSOKGLAWVLPHPRYSWKESTHADILVRLHESIOFSERTLP
CLPDSAVRLPKPTDGIAMGWSIDOGVLPHPORTOKKVPILIOSELCSLYMGAGQ
EATITGMCAGTIEBERDAGCIGDSGPTMCQVDHMLLTGLIISMEGCAADDRPVYTS
LTAHSWQRIQGVQLKRYLADSGDTGS"
BASE COUNT      280 a      375 c      372 g      294 t
ORIGIN
alignment_scores:
    Quality: 1275.50      Length: 312
    Ratio: 4.621          Gaps: 2
    Percent Similarity: 88.462      Percent Identity: 75.641
alignment_block:
US-10-040-803-7 x AB010778 ..
Align seg 1/1 to: AB010778 from: 1 to: 1321
1  MetValValSerGlyAlaProProAlaLeuGlyGlyLysLeuGlyThr 17
||||:|||||  |||:|||||  |||:|||||  |||:|||||
80  ATGATGATTCCTCCACACTCCCGCAGCAGCTGGGTGGGAGCAGTTCA 129
||||:|||||  |||:|||||  |||:|||||  |||:|||||
17  rPheThrSerLeuLeuLeuAlaSerThrAlaIleLeuAsnAlaAla 34
|||:|||||  |||:|||||  |||:|||||  |||:|||||
130  CTTAATCCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 179
||||:|||||  |||:|||||  |||:|||||  |||:|||||
34  rglleProValProProAlaCysGlyLysProGlnGlnLeuAnaArgVal 50
|||:|||||  |||:|||||  |||:|||||  |||:|||||
180  CCATCCAGATGTCCCGACAGCTGTGGAGAGCTCAGCAGCTGAACCGG 229
||||:|||||  |||:|||||  |||:|||||  |||:|||||
51  ValGlyGlyLysAspSerThrAspSerGluTrpProTrrPleValSer1 67
||||:|||||  |||:|||||  |||:|||||  |||:|||||
230  GTGGAGGTGAGAGCAGCATGATGATGCCAGTGGCCCTGATGTAGCAT 279
||||:|||||  |||:|||||  |||:|||||  |||:|||||

```

```

67 eGlnlyAsnGlyThrHisCysAlaGlySerLeuThrSerArgT 84
1  |||||
280 CCGAAGAAATGGCTCCACCACTGTGACAGGCTCCCTGCTACCAACCCGCT 329
84 rPValIleThrAlaAlaHisCysPheIysAspAsnLeuAsnIlyProIyr 100
1  |||||
330 GGGTGTCTACAGCCCGGACCTGCTTAAGACAAATATGGACAAACCACTCT 379
101 LeuPheSerValIleuGlyAlaIleTPValGluIleuGlyAsnProGlySerAr 117
1  |||||
380 CTGTTCTCAGTATGTGGGGGCTGGAAGCTGGGAGAGCCGACCCCAAG 429
117 gSerGlnIlyValGlyAlaIleTPValGluIleuProHisProValIlyrSert 134
1  |||||
430 GTCGCAAGAAAGTATGCTTGGTGGCTGCTGCTCACCACCAAGTATTCCT 479
134 rPlySgluIlyAlaCysAlaAspIleAlaLeuValArgIleuGluArgSer 150
1  |||||
480 GGAAGAGAGGAAACCCATGCAAGACATTCCTGCTGGCTGGAACACTCC 529
151 IleGlnPheSerGluArgValIleuProIleCysLeuProAspAlaSerI 167
1  |||||
530 ATCCAGTTCTCTAGGGGATTCCTGCCATCTGCTACCTGCTCTCTCT 579
167 eHisLeuProProAsnThrHisCysTrpIleSerGlyTrpGlySerIleG 184
1  |||||
580 CCGTCTCCCTCCCAAGACCGACTGCTGGATTCGCGGCTGGGGAAGCATCC 629
184 InsPglValIleProleuProHisProGlnThrLeuGlnIlySerVal 200
1  |||||
630 AGATGGAGTGGCCCTGCCCCACCTCAGACCTTACGAAGCTGAAGGAG 679
201 ProIleIleAspSerGluValIlySerHisLeuTyrTrpArgIlyAlaG 217
1  |||||
680 CCATCATTCACACTCCGAACCTGCAAAAGCTGTACTGGCGGGAGACCGG 729
217 yGlnGlyProIleThrGluAspMetLeuGlyAlaGlyTyrLeuGluGly 234
1  |||||
730 TCAGGAAGCCATCAGGAGGCGATGCTGCTGCTGCTGCTGCTGCTGCTG 779
234 IuArgAspAlaCysLeuGlyAspSerGlyGlyProleuMetCysGlnVal 250
1  |||||
780 AGCGGAGTCTTGTCTGGCGACTCTGGGGCTGCCCTGAGTGTCCAGGTG 829
251 AspGlyAlaIleuLeuAlaGlyIleIleSerTrpGlyGluGlyCysAl 267
1  |||||
830 GATGACCACTGGCTACTGACTGGCATATCAGCTGGGAGAGGCGTGC 879
267 agIuArgAsnArgProGlyValIlyrIleSerLeuSerAlaHisArgSert 284
1  |||||
880 GGAC...GACCGGCGCGGTGTGTACACACCACTCTGCTGCTGCTGCT 926
284 rPValGlnIlyIleValGlnGlyAlaIleuArgGlyArgAlaGlnGly 300
1  |||||
927 GGGTGCAGAAAGATGCTCAAGGGGTGACGCTGGCGGGTACTTGGCGGAC 976
301 GlyGlyAlaLeuArgAlaProSerGlnIlySer 311
1  |||||
977 AGTGGGAGACAGGAGAGCTCTTAATCTAGGATCT 1010
seq_name: gb_10:RA5642

```

```

REFERENCE 1 (bases 1 to 1259)
AUTHORS Lathe,R.
TITLE Direct Submission
JOURNAL Submitted (17-Apr-1998) Lathe R., Ctr. for Genome Research and Ctr.
for Neuroscience, University of Edinburgh, West Mains Road,
Edinburgh, EH9 3JG, UNITED KINGDOM
REFERENCE 2 (bases 1 to 1259)
AUTHORS Davies,B.J., Pickard,B.S., Steel,M., Morris,R.G. and Lathe,R.
TITLE Serine proteases in rodent hippocampus
JOURNAL J. Biol. Chem. 273 (36), 23004-23011 (1998)
MEDLINE 98389725
FEATURES
    source          location/Qualifiers
                    1..1259
                    /organism="Rattus rattus"
                    /strain="Fisher"
                    /db_xref="taxon:10117"
                    /clone="BSP2"
                    /issue_type="brain"
                    /tissue_type="hippocampus"
                    /dev_stage="adult"
                    <94..162
                    /gene="bsp2"
                    <94..987
                    /gene="bsp2"
                    /codon_start=1
                    /product="serine protease"
                    /protein_id="CA006644.1"
                    /db_xref="GI:3559980"
                    /db_xref="SPL:088781"
                    /translation="ARGSSPPGLTFLILPSATVSANIRGSPDCKPQQLNRVYGE
DSADAMPWIVSLTKNGSHHCASLLTNRMVYSAACFSSNMKDPVSLVIGAMKLG
NPNRSOKVIAVLVLPPIRPSREKTHADIAIVLRPPIQFERRILPCLPSSVHLF
PNTNCMIAGSGSIODGVLPPIRPIQFERRILPCLPSSVHLF
CYLEGRDADICGDSGSLMGCOVDHMLLNGIISWGSCAERNRPVYTSILAHRPVQ
RIVQGVQLRRLADSDTGS"
                    94..987
                    /gene="bsp2"
                    163..984
                    /gene="bsp2"
                    /product="serine protease"
BASE COUNT      263 a      352 c      362 g      282 t
ORIGIN
alignment_scores:
    Quality: 1223.00      Length: 299
    Ratio: 4.668          Gaps: 1
    Percent Similarity: 87.625      Percent Identity: 75.585
alignment_block:
US-10-040-803-7 x RA5642 ..
Align seg 1/1 to: RA5642 from: 1 to: 1259
5 GlyAlaProProAlaLeuGlyGlyCysLeuGlyThrPheThr.SerT 21
1  |||||
75 GGGGATCCCGACAGGAGGCGGCTTACATCACCACGAGCAGCTAACCC 124
21 euIleuLeuAlaSerThrAlaIleLeuAsnAlaAlaArgIleProVal 37
1  |||||
125 TCTTATCTCTTGGCCCTCAGCCACAGTACAGTGCAGCCAGCTCGAGGG 174
38 ProProAlaCysGlyIlyProGlnGlnIleuAsnArgValAlaGlyGly 54
1  |||||
175 TCCGCAAGATTTGGGAAGCGCAGCTGAACAGGCTGCTGGAGTGA 224
54 uAspSerThrAspSerGluIleuProProIleValAlaSerIleGlnIlyAsnG 71
1  |||||
225 GGACAGCGCGGATGCCAGTGGCTTGAATCTTATCTCAGACAGC 274
71 lYThrHisCysAlaGlySerLeuThrSerArgTrpValIleThr 87
1  |||||

```

```

275 GCTCCACACACTGCGCAGGCTCTGCTACCAACCGCTGGGTGTCTCA 324
88 AlaAlaHisCysPheIysAspAsnIleuSlnysProTyrIlePheSerVa 104
325 GCCGACACTGCTTTAGCAGCATATGACAAACCATCTCGTACTCAGT 374
104 IleuLeuGlyAlaTyrGlnIleuGlyAsnProGlySerArgSerGlnysV 121
375 ATTGTTGGGGGCGCTGGAGAGTAGAGCAACCGCCCAAGCTCAGAAAG 424
121 aIGlyValAlaTrrValGluProHisProValTyrSerTrpLysGluGly 137
425 TAGGATTTGGCTTCGGTCTGCCCAACCCAGATTTCTCGGAGAGAGGG 474
138 AlaCysAlaAspIleAlaLeuValArgLeuGlnArgSerIleGlnPheSe 154
475 ACCCATGAGACATTCGCCCTGGTGGCTGGAAGCCCATCAATCTC 524
154 rGluArgValIleuProIleCysLeuProAspAlaSerIleHisLeuPro 171
525 TGAGCGGATCCTGCCCATCTGCTACGACTCCTCGTCCATCTCCCTC 574
171 roAsnThrHisCysTrpIleSerGlyTrpGlySerIleGlnAspGlyVal 187
575 CCAACACCACTGCTGGATTCGCGGAGGAGAACATCCAGATGAGATG 624
188 ProLeuProHisProGlnThrLeuGlnIleuSlnysValProIleIleAs 204
625 CCCGCGCCCGCCCTCAGACCCCTCAGAGCTGAAGTGGCCCATCATCGA 674
204 pSerGluValCysSerHisLeuTyrTrpArgGlyAlaGlyGlnIlePro 221
675 CCTGGAACCTCTGCAAAAGTTTGTACTGGCGGAGAGCTGTCAGAGACCA 724
221 leThrGluAspMetLeuCysAlaGlyTyrLeuGlnGlyAluArgAspAla 237
725 TCACCGAGACATCTGTGTCTGTACCTGGAAGGAGCGGAGCGCT 774
238 CysLeuGlyAspSerGlyGlyProLeuMetCysGlnValAspGlyAlaTr 254
775 TGCTGGGCGCAGCTCTGGGGGGTCCCTGATGTGCCAGGTGATGACCCCTG 824
254 pLeuLeuAlaGlyIleIleSerTrpGlyGlnGlyCysAlaGluArgAsnA 271
825 GCYACTGACGGCGATATCATCAGCTGGGAGAGGCTGGCGGACGCAAC 874
271 rPProGlyValTyrIleSerLeuSerAlaHisArgSerTrpValGlnIly 287
875 GGCCGGGCGGTGTACACCAAGCTCTAGCTACCGCCCTGGGTGACAGAG 924
288 IleValGlnGlyValGlnLeuArgGlyArgAlaGlnGlyGly 302
925 ATGCTGCAGAGGGGTGAGCTGCGAGGCGCTTGGCGGACAGCTGGG 969

```

seq\_name: gb\_pr:AC003965

seq\_documentation\_block: 34642 bp DNA linear PRI 31-DEC-1997

LOCUS AC003965

DEFINITION Homo sapiens chromosome 16, cosmid clone 325D7, complete sequence.

ACCESSION AC003965

VERSION AC003965.1 GI:2734091

KEYWORDS HTG.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 34642)

AUTHORS Rieke,D.O., Bruce,D., Muntz,M., Doggett,N., Munk,C., Saunders,E., Goodwin,L., Bryant,J., Tesmer,J., Chasteen,L., Thompson,S., White,S., Ueng,S., Tatum,O., Campbell,C., Fawcett,J. and Deaven,L.

TITLE Sequencing of Human Chromosome 16p13.3

JOURNAL Unpublished

```

REFERENCE 2 (bases 1 to 34642)
AUTHORS Rieke,D.O., Wagner,R.P. and Muntz,M.O.
TITLE Large Scale Sequence Analysis and Annotation with the Sequence Comparison Analysis (SCAN) System
JOURNAL Unpublished
AUTHORS 3 (bases 1 to 34642)
Rieke,D.O., Bruce,D., Muntz,M., Doggett,N., Munk,C., Saunders,E.,
Robinson,D., Jones,M., Buckingham,J., Chasteen,L., Thompson,S.,
Goodwin,L., Bryant,J., Tesmer,J., Meinke,L., Longmire,J.,
White,S., Ueng,S., Tatum,O., Campbell,C., Fawcett,J. and Deaven,L.
TITLE Direct Submission
JOURNAL Submitted (31-DEC-1997) Center for Human Genome Studies, DOE Joint
Genome Institute, Los Alamos National Laboratory, MS M888, Los
Alamos, NM 87545, USA
COMMENT Sequence submitted by:
DOE Joint Genome Institute.
FEATURES
source
1..34642
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="16p13.3"
/clone="325D7"
/complement(96..381)
/rpt_family="Alu"
/complement(955..1255)
/rpt_family="Alu"
/complement(1803..2246)
/rpt_family="Alu"
2626..2809
/rpt_family="Alu"
3105..3373
/rpt_family="Alu"
4249..4384
/rpt_family="Alu"
/complement(4506..4797)
/rpt_family="Alu"
6206..6510
/rpt_family="Alu"
/complement(6664..6751)
/rpt_family="Alu"
/complement(8846..9171)
/rpt_family="Alu"
9425..9509
/rpt_family="Alu"
9509..9573
/rpt_family="MIR1b"
9672..9766
/rpt_family="MSTC"
10069..10364
/rpt_family="MSTC"
10370..10674
/rpt_family="Alu"
/rpt_family="Alu"
/complement(10884..11179)
/rpt_family="Alu"
11211..11522
/rpt_family="Alu"
11493..11799
/rpt_family="Alu"
/rpt_family="Alu"
/complement(11793..12160)
/rpt_family="L1"
13964..14098
/rpt_family="L1"
/rpt_family="MIR"
/complement(14456..14468)
/rpt_family="Alu"
16240..16446
/feature="45% Swiss-Prot.P18127"
/join(<17398..17540,17771..18048,19600..19757,20293..20878)
/feature="SP001LA"
/join(<17398..17540,17771..18048,19600..19757,20293..20529)
/feature="SP001LA"
/feature="putative serine protease"
/codon_start=1

```

```

/evidence=not_experimental
/product="SP001LA"
/protein_id="AAB93671.1"
/db_xref="GI:2734092"
/translation="LNRYVGGSDSTDEMPYIVSIQKNGTHHCAGSILTSRWVITAAH
CEADNLNRPYLFVSLGAWQDGNPSRSOKGVANVEHPHYSMKEGACADIALYRL
RSIQSERVLPICLPDASIHLPNTHCWISGWSIQDGVLPFPQTLQKLVITDSE
VCSHLWRGAGGPIEDMTCAGYLEGERDACLDSGVLKQVDMGVAMLAITISWG
GCAERNRPGVYISLSAHRSMVEKIVGVQLRGRAGGGGALRAPSGSGAAARS"
repeat_region
/rpt_family="Alu"
complement(18257..18556)
polya_site
20873..20878
/feature="SP001LA"
complement(22028..22333)
repeat_region
/rpt_family="Alu"
complement(22376..22722)
repeat_region
/rpt_family="Alu"
complement(24599..25217)
repeat_region
/rpt_family="Alu"
complement(25525..25603)
repeat_region
/rpt_family="Alu"
complement(25703..26005)
repeat_region
/rpt_family="Alu"
complement(26818..26888)
repeat_region
/rpt_family="LTR8"
complement(26927..27263)
repeat_region
/rpt_family="MER4"
27294..27592
/rpt_family="Alu"
complement(27650..27931)
repeat_region
/rpt_family="MER4b"
complement(29107..29227)
repeat_region
/rpt_family="Alu"
complement(30340..30649)
repeat_region
/rpt_family="Alu"
33294..33459
/feature="Serrine Protease SP002LA catalytic exon A"
32602..332780
/feature="Serrine Protease SP002LA catalytic exon B"
33314..33474
/feature="Serrine Protease SP002LA catalytic exon C"
33691..33966
/feature="Serrine Protease SP002LA catalytic exon D"
BASE COUNT 7787 a 9148 c 8812 g 8895 t
ORIGIN
...
Percent Scores:
Quality: 1094.50 Length: 1055
Ratio: 3.895 Gaps: 5
Percent Similarity: 26.635 Percent Identity: 26.635
alignment_block:
US-10-040-803-7 x AC003965 ..
Align seg 1/1 to: AC003965 from: 1 to: 34642
36 ProValProProAlaCysGlyLysProGlnGlnLeuAsnArgValValG1 52
|||||
17365 CCAGTTCGCCACGCTGTGGAGAGCCCGACGAGCTAAACGGGTTGTGG 17414
|||||
52 YGlyGluAspSerThrAspSerGluTrpProTrpPileValSerIleGlnL 69
|||||
17415 CGGCGAGGACAGACTGACAGCGAGTGGCCCTGATCGTGAAGATCCAGA 17464
|||||
69 yAsnGlnGlyThrHisLysCysAlaGlySerLeuLeuThrSerArgTrpVal 85
|||||
17465 AGAATGGGAGCCACCACTGGCGAGTTCTGTCTACCAAGCGCTGGGTG 17514
|||||
86 IleThrAlaAlaHisCysPheLys..... 93
|||||
17515 ATCACTGCTGCCACCTGTTCAAGAGATATGTAAGCGCGCTGGGGCAC 17564
|||||

```

```

93 ..... 93
17565 TTGATCTTCTAAGGCCCTGAGGCTTGTGTGCCAGAGCCCTGGGCTCCTT 17614
93 ..... 93
17615 ACAGGCTCTGCTTCGGGGCCTCTGTTCAATCTTGTGCCCCGTGATTTCTA 17664
93 ..... 93
17665 CCTAGAAGCCTCTCTGCTTTCAGGGGCTTTCACACCATGCTTTCC 17714
93 ..... 93
17715 CAGGTACCTCTCCAGACGAGGCTGGCTGATCTCCCTTCCCTTCC 17764
94 ..Asp..AsnLeuAsnLysProTyrLeuPheSerValLeuGlyAlaT 109
|||||
17765 TTCCAGCAACCTGAAACAAACCATACCTGTTCTCTGCTGCTGGGGGCT 17814
109 rPcInLeuGlyAsnProGlySerArgSerGlnLysValGlyValAlaT 125
|||||
17815 GCGAGCTGGGGAAACCTGGCTCTGGTCCAGAGAGTGGGTTCCTGG 17864
126 ValGluProHisProValTyrSerTrpLysGluGlyAlaCysAlaSp 142
|||||
17865 GTGAGGCCCAACCTGCTGATATCTCTGGAAGAGTGCCTGTGACAGAT 17914
142 eAlaLeuValArgLeuGluArgSerIleGlnPheSerGluArgValLeu 159
|||||
17915 TGCCCTGCTGCTCTCGAGGCTCCATACAGATTCTGAGCGGGTCTTGC 17964
159 rLleCysLeuProAspAlaSerIleHisLeuProPheAsnTrpHisCys 175
|||||
17965 CCATCTGCTACCTGATGCTCTTATCCACTCTCCCAACACCATCTGC 18014
176 TrpIleSerGlyTrpGlySerIleGlnAspGly..... 186
|||||
18015 TGGATCTCAGGCTGGGGAGCATCCAGATGAGGTGCGCTTCTTCTTC 18064
186 ..... 186
18065 TCCAAAGTGTGGGGGAGATGGGAGACAGAGATGCGACACTCAGG 18114
186 ..... 186
18115 ACACAGACCATGACGACAAAGGTGGAGGCCAGAGCCACAGCATGAC 18164
186 ..... 186
18165 AGAGGTGGAAGAAAGATGGGCTGCAATCCAGAAACCTTGGTGGCA 18214
186 ..... 186
18215 ATTTCAGGCATTTCTCTTTCTTTCTTTCTTTCTTTCTTTT 18264
186 ..... 186
18265 TTTTCTTTGAGACAGGCTGTACTGTGTCACCGAGCTGAGTGCAGT 18314
186 ..... 186
18315 GGTGTGATCTAATCACTGAGGCTTTGCCCTCCAGGCTCAAGTATCC 18364
186 ..... 186
18365 TCCCACTCAGGCTCCAGAGTACGTGAAGTGTGTGTACACCA 18414
186 ..... 186
18415 CTTGCTAATTTTGTATTTTGTAGAGACAGGTTTCTGTGTGGC 18464
186 ..... 186

```

18465 TGGCTGCTTGAAGTGTGGCTCAAGCAATCCACCGCTCAGCCTC 18514  
186 ..... 186  
18515 CCAAGTCCGGGATTACAGGTGACGCCACTGCTGCTCATTCT 18564  
186 ..... 186  
18565 ACTCAAGATCTGCTATGTTTGAGGCAACAGCCAGACCCGTGAAG 18614  
186 ..... 186  
18615 CTATTCAATTCATATACAGAAATGCTGAAGGGGTGTGTGGAG 18664  
186 ..... 186  
18665 TAAGCTAATACAGGCTGAATATTAAGAGCCCTGATTCAGAGTGGAA 18714  
186 ..... 186  
18765 GAAATGACACCACTTAATTGTAAATATTAATTTCAACCACCACTGG 18814  
186 ..... 186  
18815 AGGTTGGAAAATCATCTCAAACTCAAACTAGATTACATTAAGCAAG 18864  
186 ..... 186  
18865 AACGAGGAGACAGCAAGAAAGTAGATGACAGATCGACAGGCAGATG 18914  
186 ..... 186  
18915 GATGATAGAGACTGGGCCAAGAAAACATACAGGAGAAATGAGGA 18964  
186 ..... 186  
18965 TACAATGAAGACCTGGAAGAAAGTTTGAAGAGAGAGCTAGAGACTTA 19014  
186 ..... 186  
19015 GTATATAAAGTAGTAGACAGATGAATAATTACAGATTCAGAAATA 19064  
186 ..... 186  
65 TTAAGAGAAATACACTACTCAGGTGTGTGGGATCCAGTAACCTGA 19114  
186 ..... 186  
19115 GGTTCACCTCTCCATGATCTGTGAGGGGCTTCTCAAGAGTAGAAT 19164  
186 ..... 186  
19165 CGCGGACTTTCAGTCCACATCTGACACACAGTGTAGACAGCTCC 19214  
186 ..... 186  
19215 AGGCAAGGCTGGGATCCCTCAGGTCTGTGTAATTTCACATCGAGTG 19264  
186 ..... 186  
19265 GACATTTATTGAACACTTCTTATAGCAGTCCAGTGTGTGATGAAGTA 19314  
186 ..... 186  
19315 ATGCAATCTGAGACAGGAGGCTTAACCATCTGTGATATTAAGAAA 19364  
186 ..... 186

19365 ATTATTAAAGAGTAGGCTTTAGGTTTCATTCATCTGTCTAATCAG 19414  
186 ..... 186  
19415 CAGGGCCAGAGGTGAGACCTGGAATTTCATTTCTGAAGCATCTTG 19464  
186 ..... 186  
19465 GAGAGACTGTAGACTAATCTGTAGCCTTTGAGAACCCCTGCTGAG 19514  
186 ..... 186  
19515 GAATGCTGTGGGGGAATAGTCTTTGAGAGGCAACGAGAGGGAG 19564  
187 ..... 187  
19565 ACCCCCAACCTGATTCGTCTCTCTCCCTCCACATTCCTGCCCCAC 19614  
192 roglThrLeuGlnIleuValProIleIleAspSerGluValCys 208  
19615 CTCACACCCCTGAGAGAGCTGAAGGTTCTATCATCTGACTCGGAATCTGC 19664  
209 SerHisLeuTyrTrpArgGlyAlaGlyGlnGlyProIleThrGluAspMe 225  
19665 AGCCATCTGTACTGGCGGGAGACAGAGACAGGACCATCCTGAGGACAT 19714  
225 tLeuGlyAlaGlyTyrLeuGlnGlyGluArgAspAlaCysLeu..... 239  
19715 GCTGTGCCGCTACTCTGAGGGGAGCGGATGCTGTGTGTGAGCT 19764  
239 ..... 239  
19765 CCTTGAGCCCCCCCCACCCCTGGCCAGAGGGCTTCGGAAAGAGACCA 19814  
239 ..... 239  
19815 TCAGTCTGTCCCACTGAGCCCAACAACCTCTCCTGGCTCTCCCA 19864  
239 ..... 239  
19865 GAAGTGGCTGTGCTGACAGTCCGGGCTCCCATAGAACAGCCTCAGCC 19914  
239 ..... 239  
19915 TGGCTCAGCCTCCTATTTGTTCAAGACATTAACCTGGGCATCCAGC 19964  
239 ..... 239  
19965 TGACACACCTGCCCTCCGAGACTTTCATTTCTTCAGGGGCTTCCA 20014  
239 ..... 239  
20015 CGCCTCAGTCCGACAGACAGCTGACTCAGGGGCTGTCTGAAGACGG 20064  
239 ..... 239  
20065 TCAGATGTTGAGGGGAGACACCGCGCTGAGAGTCTTAAGGCTGGCT 20114  
239 ..... 239  
20115 GCAGCTCTACTCTGCGCGGGGACTTCTACCTGGGTGAGAGCTCAGGT 20164  
239 ..... 239  
20165 CTTGTCTCTCAGAGAGAGAAAGGGCTCCCGGGAAGGGCTCATCTG 20214  
239 ..... 239  
20215 CCCCCTGACAGGGGAGCGGCTGATTTGACACCCCTGTGCTCACAGGG 20264  
240 ..... 240  
20265 GGTTCCTAACCTCCTCTCTCTGACAGGGGAGCTCCGGGGGCCCTCA 20314



```

247 etCysGlnValaspGlyAlaTrpLeuLeuAlaGlyIleIleSerTrpGly 263
|||||
20315 TGTGGCAGGTGGAGCGCGCTGCTGCTGGCCGCATCATCAGCTGGGGC 20364
264 GIUGIYCysAlaGluAraAsnArpProGlyValTyrIleSerLeuSerAl 280
|||||
20365 GAGGGCTGTGGAGGAGGACAGCGCCGGGCTTCATCAGCTCTGCGC 20414
280 ahIsArpSerTrpValGluTyrIleValGlnGlyValGlnLeuArpGlyA 297
|||||
20415 GCACGCTCTCTGGGTGGAGAGATCGTGCAGAGGGTGCAGCTCCGGGAC 20464
297 rgaAlaGlnGlyGlyAlaLeuArpGlnProSerGlnGlySerGlyAla 313
|||||
20465 GCGCTCAGGGGGGTGGGGCTCAGAGGACCGAGCGAGGGCTGTGGGGC 20514
314 AlaAlaArpSer 317
|||||
315 GCCGCGCGCTCC 20526
name: gb_pr: HSA306593
seq documentation block:
LOCUS HSA306593 1013 bp mRNA linear PRI 31-MAR-2001
DEFINITION Homo sapiens mRNA for marapsin (MPN gene).
ACCESSION AJ306593
VERSION AJ306593.1 GI:13516325
KEYWORDS marapsin; MPN gene.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS 1 (bases 1 to 1013)
TITLE Fortunato, M., Dando, P. M., Rawlings, N. D. and Barrett, A. J.
JOURNAL Cloning, sequencing and expression of marapsin, a human serine
proteinase
REFERENCE 2 (bases 1 to 1013)
AUTHORS Fortunato, M.
TITLE Direct Submission
JOURNAL Submitted (29-MAR-2001) Fortunato M., MRC Molecular Enzymology
Laboratory, Babraham Institute, Babraham Hall, Babraham, Cambridge,
CB2 4AF, UNITED KINGDOM
COMMENT Related entry AAI70323
FEATURES
source
location/Qualifiers
1..1013
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE 2698030"
20..892
/gene="MPN"
20..892
/gene="MPN"
/gene="MPN"
/function="putative trypsin-like activity"
/codon_start=1
/product="marapsin"
/protein_id="CAC35467.1"
/db_xref="GI:13516326"
/translation="MRRPAAYPLLLLCFSORAKATACRPRLNRMNGGOTQES
EMPQVVSIRNGSHFCGSLIAEOMVLTAAHCFRTSTSLYVLLGRQLVQPHNA
MYARVQESNPLVGYTASADVALVELAEVPTNYILPYCLDPSVTFEGNCAWY
TQWSPSEEDLPEPRITOKLAVPIIDPKCNLYSKDTEGYGPKTIKNMLCAGFE
EGKDKAGDGGGGLVCLVGSQWLDAGVISWEGECARQNRPGVYIRVTAHHNHTRIT
PKLQROPRLRGOK"
BASE COUNT 192 a 334 c 321 g 166 t
ORIGIN
alignment_scores:
Quality: 739.00 Length: 285
Ratio: 3.421 Gaps: 4
Percent Similarity: 75.779 Percent Identity: 48.772

```

```

alignment block:
US-10-040-803-7 x HSA306593
Align seg 1/1 to: HSA306593 from: 1 to: 1013
21 LeuLeuLeuAlaSerThrAlaIleLeuAsnAlaAlaArgIleProVa 37
|||||
47 CTGCTGCTGCTGCT .....TTTGGTCTCAGAGGCGCCAGAGC 84
37 1ProProAlaCysGlyIlyProGlnGlnLeuAsnArgValAlaGlyGly 54
: |||||
85 AGCAACAGCTGTGTGTCGCCAGAGATCGTAACAGATGGTGGCGGC 134
54 LuAspSerThrAspSerGluTrpProTyrIleValSerIleGlnLysAsn 70
: |||||
135 AGGACACGACGAGAGGCGAGTGGCCCTGGCAGTCAATCAGCGGAC 184
71 GlyThrHisCysAlaGlySerLeuLeuThrSerArgTrpValIleTh 87
|||||
185 GGAAGCCACTCTGCGGGGCGACCTCATCGGAGCAGTGGCTCTGAC 234
87 rAlaAlaHisCysPheLysAspAsnLeuAsnLysProTyrLeuPheSerV 104
|||||
235 GGTGGCGCAGCTGCTCCG...AACCTCGAGACGTCCTGTGACAGG 281
104 allLeuGlyAlaTrpGlnLeuGlyAsnProGlySerArgSerGlnLys 120
|||||
282 TCTGCTGGGGGCAAGGAGCTAGTGCAGCGGAGCACACAGCTATGAT 331
121 ValGlyValAlaTrpValGluProHisProValTyrSerTrpLysGlu 137
: |||||
332 GCCCGGTGAGGAGGAGTGTGAGACGACACCCCTGTACAG...GGCAGCGC 378
137 yAlaCysAlaAspIleAlaLeuValArgLeuGlnArgSerIleGlnPhe 154
: |||||
379 CTCAGCGCTGACGTGGCCGTGGAGAGTGTGAGGAGGACCACTGCTTCA 428
154 ergLuarValLeuProIleCysLeuProAspAlaSerIleHisLeuPro 170
: |||||
429 CCAATTACATCTCCCGGTGGCTGCTGACCCCTGATGATCTAG 478
171 ProAsnThrHisCysTrpIleSerGlyTrpGlySerIleGlnAspGlyVa 187
: |||||
479 ACGGCGATGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 528
187 lProLeuProHisProGlnThrLeuGlnLysLeuLysValProIleLea 204
|||||
529 CCTCTGCGCGAGACCGGATCTCTCAGAACTCGCTGTGCCATCATCG 578
204 spSerGluValCysSerHisLeuTyrTrpArgGlyAlaGlyGlnGly... 219
: |||||
579 ACACACCCAGTGAACCTGCTCTACAGCAAGACAGCAGATTGGCTTAC 628
220 .....ProIleThrGluAspMetLeuCysAlaGlyTyrLeuGlnG 233
|||||
629 CAACCCAAAACCATCAGATGATGATGCTGCTGCGGCTTGGAGAGAG 678
233 yGluArgAspAlaCysLeuGlyAspSerGlyIlyProLeuMetCysGln 250
: |||||
679 CAAGAAGATGCTCTGCAAGGGGAGCTGGGGCGCCCTGGTGCCTG 728
250 aAspGlyValTrpLeuLeuAlaGlyIleIleSerTrpGlyGlnLys 266
: |||||
729 TGGGTGAGTGTGGCTGAGCGGCGGGGTGATACCTGGGGGTGAGGGCTGT 778
267 AlaGluArgAsnArgProGlyValTyrIleSerLeuSerAlaHisArgSe 283
: |||||
779 GCCCGCAGAACCGCCAGGTGTCTACATCGTGTACCGCCACACACAA 828
283 rTrpValGluLysIleValGlnGlyValGlnLeuArgGlyArgAlaGln 300
: |||||
829 CTGATCATCGATCATCCCAAACTGCAGATTCCAGCGAGGAGTTGG 878

```

300 1ygly 301  
|||||  
879 GCGGC 883

seq\_name: gb\_ov:AB018694

seq\_documentation\_block:

LOCUS AB018694 2078 bp mRNA linear VRT 05-OCT-1999

DEFINITION Xenopus laevis xepsin mRNA for epidermis specific serine protease,

complete cds.

ACCESSION AB018694

VERSION AB018694.1 GI:6009514

KEYWORDS epidermis specific serine protease; xepsin.

SOURCE Xenopus laevis neurula cDNA to mRNA.

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;

Xenopodinae; Xenopus.

1 (sites)

Yamada, K.

The expression control of xepsin by non-axial and planar

posteriorizing signals in Xenopus epidermis

Unpublished (1998)

2 (bases 1 to 2078)

Yamada, K., Takeshima, K. and Takabatake, T.

Direct Submission

Submitted (15-OCT-1998) Kazuto Yamada, Graduate School of Human

Informatics, Nagoya University; Furo-cho, Chikusa-ku, Nagoya

464-8601, Japan (E-mail: yamada@info.human.nagoya-u.ac.jp,

Tel:+81-52-789-2572, Fax:+81-52-789-2567)

Location/Qualifiers

1. 2078

/organism="Xenopus laevis"

/db\_xref="taxon:8355"

/dev\_stage="neurula"

35. 1204

/gene="Xepsin"

/note="serine protease"

/codon\_start=1

/product="epidermis specific serine protease"

/protein\_id="BA84941.1"

/db\_xref="GI:6009515"

/translation="MLQYLSFVLIFFHHOAGCVPIVSNRIYGVGSKRGEMPMQISL

YKSDSICGSLDLSWMTAAHACIDSLDVSYYTYIAYOIASAPDNGSTVSRGYSTIK

HPDQYESSGSDIALIELEKPVPTPTPLPICPSOYDPAAGTMCWTGNGNIQEGT

PLISPKTQKAEVAILIDSSVCGTMTESLGIPTDFSFIDEDMVCAGKEGIDACQGD

SGGLVCNVNVMQLGLIVSWGCAEENRPGVYTKVQYQDMKTWPLIVSEEBP

SVAPISGSIAPSGPSLIGRGSVASTISQTEAGSVNSIETDKNSTTIFETEMSMS

NNTMNEFFSVLSTISTALRINETKTIDNEAQIHACSLHTIALTLILYLFIRFV"

BASE COUNT 603 a 455 c 437 g 583 t

ORIGIN

alignment\_scores: Quality: 661.50 Length: 282

Ratio: 3.227 Gaps: 7

Percent Similarity: 72.695 Percent Identity: 47.163

alignment\_block: US-10-040-803-7 x AB018694 ..

Align seg 1/1 to: AB018694 from: 1 to: 2078

38 PROPRo.AlaCysGlyLysProGlnGlnLeuAsnArgValAlaGlyGlyG 54  
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
73 CCACCAAGCCTGIGGTACCAAGTAATCAATGCGATGTTGGGGGCA 122  
54 laaspeThrasperGluTrpProTrrPleValSerIleGlnLysasn 70  
||||| ::||| ||||| ||||| ::||| ::||| ::||| ::|||  
123 TGCACAGTAAAGAGGGGGAATGGCTTGCAATAGACCTGAGCTACAAA 172

71 GlyThrHisCysAlaGlySerIleuLeuThrSerArgTrpValIleTh 87  
::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

173 AGTAACTTATCTGCGGGGAGTGGCTTTACAGACAGCTGGGTGATGAC 222

87 fAlaIleHisCysPheLysAspAsnLeuAsnLysProTrrPleValIleTh 104

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

223 TGTGCTCAGTCGATTT...GATTCGTTGGATGTTTCATAC...TATACG 266

104 albeuLeuGlyAlaTrpGlnLeuGlyAsnProGlySerArgSerLys 120

|| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

267 TTATCTCGGTGATATACAGCTCTGTCGCCCGACAAATCTTCACAGATGC 316

121 ValGlyValAlaTrpValGluProHisProValTrrSerTrpLysGlu 137

||||| ::||| ||||| ||||| ||||| ||||| ||||| |||||

317 AGAGGAGTGAAGAAAGTAAAGCAAGCCAGATTCATAC...GAAGG 363

137 yAlaCysAlaAspIleAlaLeuValArgLeuGlnArgSerIleGlnPhe 154

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

364 ATCCAGTGGTGCATTTGCTTATTAAGTGGAGAAACCTGTCACCTTTTA 413

154 erGluArgValLeuProLleCysLeuProAspAlaSerIleHisLeuPro 170

::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

414 CCCTTACATCCTACACCATGCTCCCTCCCTCAGAGATTCATTTGCA 463

171 ProAsnThrHisCysTrpLysSerGlyTrpGlySerIleGlnAspGly 187

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

464 GGAGGAGTATGTGCTGGTGACTGATGGGAAATATTCAGAAAGAGAAC 513

187 lProLeuProHisProGlnThrLeuGlnLysLeuValProIleIleA 204

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

514 CCCTTTAAACAGCCCTAAACATATTCAGAGGACAAAGTTCGCATCATAG 563

204 sPseSerGluValCysSerHisLeuTrrTPArgGlyAlaGlyGlnGlyPro 220

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

564 ACAGTTCACTGTGTGGACCATGTACCAATCCAGCTGGCTACATCCCT 613

221 .....lleThrGluAspMetLeuCysAlaGlyTrrLeuGlnGlu 233

|| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

614 GATTCACGCTTCATTCAGAGGACATGCTGTGCTGCTTCAAGGAGGG 663

233 yGluArgAspAlaCysLeuGlyAspSerGlyGlyProLeuMetCysGln 250

::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

664 ACGTATGATGCTGCTCCAGGGGACACTGAGTGCACCTCTTGATGATATG 713

250 alAspGlyAlaTrpLeuLeuAlaGlyIleIleSerTrpGlyGlnGlyCys 266

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

714 TCACAAATGTGTGCTTCAGCTCGGTATGTGAGCTGGGATATGGCTGC 763

267 lAlaGluArgAsnArgProGlyValTrrLysSerLeuSerAlaHisArgse 283

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

764 GCAGAACCAATAGACCTGCTGTCTATACCAAGTTCAGTTCACCGAGA 813

283 TTrpValGluLysIleValGlnGlyValGlnLeuArgGlyArgAlaGln 300

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

814 TTGCTTGAAG...ACCAATGTGCTTTPATTTGTTGTCAGTGAG 854

300 1yglyGlyAlaLeuArgAlaProSerGlnGlySerGlyAlaAla 314

||||| ::||| ||||| ||||| ||||| ||||| ||||| |||||

855 AAGGACCAAGTGT...GCACCAAGCATTTGGACCAAGCATTTGCA 895

seq\_name: gb\_pat:AX098193

seq\_documentation\_block:

LOCUS AX098193 1796 bp DNA linear PAT 30-MAR-2001

DEFINITION Sequence 105 from Patent WO0118542.

ACCESSION AX098193

VERSION AX098193.1 GI:13515276

KEYWORDS

SOURCE

ORGANISM

human.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

```

REFERENCE 1 (bases 1 to 1796)
AUTHORS Lee,J., Thompson,P. and Lillie,J.
TITLE Identification, assessment, prevention, and therapy of ovarian
JOURNAL cancer
Patent: WO 0118542-A 105 15-MAR-2001;
Millennium Predictive Medicine, Inc. (US)
FEATURES
Source
1. 1796
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 327 a 599 c 511 g 359 t
ORIGIN
alignment_scores:
Quality: 607.50 Length: 292
Ratio: 2.993 Gaps: 7
Percent Similarity: 69.521 Percent Identity: 44.178
seq_block:
-040-803-7 x AX098193 ..
Align seg 1/1 to: AX098193 from: 1 to: 1796
10 LeuGlyGlyGlyCysLeuGlyThrPheThrSerLeu.....Le 23
||||| ||| ||||| : : : |||||
170 CTGGGGCTGCGACCTGGGGCTGCGCATTCCTGCTATCTTGATTT 219
23 uleuLaserThrAlaLeuAsnAlaAlaArgLleProValProPro 40
||||| ||| : : : ||||| : : : |||
220 ACTCCGGCTCGGGACAGAGCGGAGGCGGAGAACCTCC..... 259
40 lacysGlyLysProGlnLeuAsnArgValValGlyGlyLysPse 56
||||| ||| : : : ||||| : : : |||
260 ..TGGGGTGTGGCCCCCAAGCA...CCCATCAGAGTGGCAGAGTGA 304
57 ThrAspSerGluTrrProTrrPleValSerLleGlnLysAsnGly 73
: : : : : ||||| ||||| : : : : :
305 GTCCGGCTGACGTGGCTGCTGAGTGCATCATCATGAGGAGGCTCA 354
73 HisCysAlaGlySerLeuLeuThrSerArgTrrValLleThrAla 90
| ||||| ||||| : : : ||||| : : : |||
335 TGTGTGTGTGTGCTCTCTGCTGCTGAGCAGTGGTGTCTGACGTCTC 404
90 lscysPheLysAspAsnLeuAsnLysProTrrLysPheSerValLeu 106
||||| : : : : : ||||| : : : |||
405 ACTGTTCTCCAGCGAGCAACCAAGAA...GCCTATGAGTCAAGCTG 451
107 GAlaLtrPglInLeuGlyAsnProGlySerArgSerGlnLysValG 123
||||| ||||| : : : ||||| : : : |||
452 GGGGGCCACACAGTACCTCTACTCTGAGGAGGAGGAGGAGGAGCT 501
123 lAlaLtrPglInProHisProValTrrSerTrrLysGlnLysAla 140
: : : ||||| ||||| : : : ||||| : : : |||
502 GAAGAGCATCATCCCCACCCAGCTAC...CTCCAGAGAGGCTCCCG 548
140 lAspPleAlaLeuValArgLeuGlnArgSerLleGlnPheSerGln 156
: : : ||||| : : : ||||| : : : |||||
549 GCGACATGACCTCTCCACACTGACAGACCACTCATCTCCCGCTAC 598
157 ValLeuProLleCysLeuProAspAlaSerLleHisLeuProPro 173
: : : ||||| ||||| : : : ||||| : : : |||
599 ATCCGCCCATCTGCTCTCCAGCAACGAGCTCTCCCAAGGCGCT 648
173 HisCysTrpLieserGlyTrrPglSerLleGlnAspGlyLysPle 190
||||| : : : ||||| : : : ||||| : : : |||
649 CCATGCACTGCTGCTGCTGGGTGCTGCTGCTGCTGCTGCTGCTGCT 698
190 rHisProGlnThrLeuGlnLysLeuValProLleLysAspSerGln 206
||||| : : : ||||| : : : ||||| : : : |||
699 TGACCCCAAGGCACTGACAGCAACTGAGTGTCTGATGATGCTGAG 748
207 ValCysSerHisLeuTrrPargLysAla...GlyGlnGlyPro.... 220

```

---

```

749 ACCTGTAATGCTGCTGACAAACATGACGAGCAAGCCGAGAGCCGACTT 798
221 lIerhGluAspMetLeuCysAlaGlyTrrLeuGlnGlyLysAlaG 237
: : : ||||| : : : ||||| : : : |||||
799 TGTCCAAAGAGACATGCTGTGCTGCTGCTGCTGCTGCTGCTGCTG 848
237 lacysLeuGlyAspSerGlyGlyProLeuMetCysGlnAlaSpGly 253
||||| ||||| ||||| ||||| ||||| |||||
849 CCTGCGAGGTGACTGCTGGGGCCACTCTCCCTCTGAGAGGCTTC 898
254 TrpLeuLeuAlaGlyLleLieserTrrPglLysGlyCysAlaGlu 270
||| ||||| : : : ||||| : : : |||||
899 TGCTACTGACGCGCATTTGAGACTGGGAGAGATGCTGTGGGCGCCGA 948
270 nArgProGlyValTrrLieserLeuSerAlaHisArgSerTrrPgl 287
||||| ||||| ||||| ||||| ||||| |||||
949 CAGCGCTGTGTGTACACTGTGCTGCTGCTGCTGCTGCTGCTGCTG 998
287 yslIeValGlnGlyValGlnLeuArg 295
: : : ||||| |||||
999 GCAAGGTGACAGAACTCCAGCTCTGT 1024
seq_name: gb_pr:BC001462
seq_documentation_block:
LOCUS BC001462 1809 bp mRNA linear PRI 12-JUL-2001
DEFINITION Homo sapiens, prostate, 8 (prolactin), clone MGC:2133
IMAGE:3138532, mRNA, complete cds.
ACCESSION BC001462
VERSION BC001462.1 GI:12655206
KEYWORDS MGC.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 1809)
AUTHORS Strausberg,R.
TITLE Direct Submission
JOURNAL Submitted (12-DEC-2000) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
REMARK NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT Contact: MGC help desk
Email: cgaabs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: National Institutes of Health Intramural,
Sequencing Center (NISC),
Gaithersburg, Maryland;
Contact: http://www.nisc.nih.gov/
Web site: nisc.mgc@nih.gov
Sherchenko,Y., Wetherby,K.D., Beckstrom-Sternberg,S.M.,
Benjamin,B., Blakesley,R.W., Bouffard,G.G., Brinkley,C., Brooks,S.,
Dietrich,N.L., Guan,X., Gupta,J., Ho,S.-L., Karlins,E., Legaspi,R.,
Lim,M., Maduro,Q.L., Mastello,C., Mastrian,S.D., McCloskey,J.C.,
McDowell,J., Pearson,R., Snyder,B., Stantiripop,S., Thomas,P.J.,
Tiongson,E.E., Touchman,J.W., Tsurgeon,C., Vogt,J.L., Walker,M.A.,
Zhang,L.-H. and Green,E.D.
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IMAGE Plate: 4 Row: 3 Column: 3
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 1143193.
FEATURES
Source
1. 1809
/organism="Homo sapiens"
/db_xref="locusID:5652"
/db_xref="taxon:9606"
/clone="MGC:2133 IMAGE:3138532"

```

CDs

/tissue\_type="Placenta, choriocarcinoma"  
 /clone\_id="NH\_MGC\_21"  
 /lab\_host="DH10B-R"  
 /note="Vector: POTB7"  
 185..1216  
 /codon\_start=1  
 /product="protease, serine, 8 (prolactin)"  
 /protein\_id="AA01462.1"  
 /db\_xref="GI:12655207"  
 /translation="MAKGVILPGGLGAVAILLYLGLRSGTGAEGAEAPCGVAPQAR  
 ITGSSAVAGOMPMQVSYITPEYVHCGGLSVDSQWVLSAHCPSHHKAEYVKLGA  
 HQLDSYSEDARVSTLKDILPHPSYLDGSGSDYALQLSPTFSRYTIPICIPANA  
 SPNGSLHCTVYGMHVAIPSVSLTPKPLQLEPLLSRETCLNLTIDAKPEEPHYQ  
 EDWVCAQYVGGDADCGDGGPLSCPEVILWLTGIVSGDAGARNRPGVYTLAS  
 YASWIOSKYTELQPRVYPTQESQPDSPNLCGSHLAFSSAPADQILRPILFLPLGLALG  
 LSPWLVSEH"

BASE COUNT 322 a 605 c 517 g 365 t  
 ORIGIN

ent\_scores:  
 Quality: 607.50 Length: 292  
 Ratio: 2.993 Gaps: 7  
 Percent Similarity: 69.321 Percent Identity: 44.178

alignment\_block:  
 US-10-040-803-7 x BC001462 ..

Align seg 1/1 to: BC001462 from: 1 to: 1809

```

10 LeuGlyGlyGlyCysLeuGlyThrPheThrSerLeu.....Le 23
||||| ||| |||||::: ||||| ||
203 CTGGGGCTGGGCGAGCTGGGGCTGGCCATTGCTTATCTTGGATT 252
23 uLeuAlaSerThrAlaIleLeuAsnAlaIlaArgIleProValProPro 40
||||| ||| :::::|||||::: |||
253 ACTCGGCTGGGACGAGCGGAGGGGACGAGAGCTCC..... 292
40 lAcysGlyLysProGlnGlnLeuAsnArgValValGlyGlyLysAspSer 56
||||| ||| :::::|||||::: |||
293 ..TGCGGTGGGGCCCCCAAGCA..CGCATCAGAGTGGCAGAGTGA 337
57 ThrAspSerGluTrpProTrpIleValSerIleGlnLysAsnGlyThrH 73
::: :::::||||| ||||| |||
338 GTGCGCGGTGAGTGGCCCTGGCGAGTCAACATCATTGAGGAGCGCTCA 387
73 sHisCysAlaGlySerLeuLeuThrSerArgTrpValIleThrAlaIlaH 90
| ||||| ||||| |||||::: ||||| |||||
98 TGGTGTGGTGGCTCTCTGCTGTGAGCAGTGGGTGTGCTGAGCTGCTC 437
90 lscysPheLysAspAsnLeuAsnLysProTyrLeuPheSerValLeuLeu 106
||||| ||| ::::: ||||| |||
438 ACTGCTTCCCGCCAGCAGCAGCAGCAAGAA..GCCATGAGGTCAAGCTG 484
107 GlyAlaTrpGlnLeuGlnAsnProGlySerArgSerGlnLysValGlyVal 123
||||| ||||| ::: ::::: |||
485 GGGGCCACACAGCTAGACTCTCTACTCCGAGGAGCCAGGTCAGACCTT 534
123 lAlaTrpValGluProHisProValTyrSerTrpLysGlnGlyAlaCysA 140
: ||||| ||||| |||
535 GAAGGACATATCCCGCCAGCTAC...CTCCAGGAGGCTCCAGG 581
140 lAspIleAlaLeuValArgLeuGluArgSerIleGlnPheSerGluArg 156
::: ||||| |||||::: ||||| |||
582 GGCACATGACTCTCCAACTCAGCAGCAGCCATCACTTCTCCCGCTAC 631
157 ValLeuProIleCysLeuProAspAlaSerIleHisLeuProProAsnTh 173
::: ||||| ||||| |||
632 ATCCGGCCATCTGCTCCCTGACGCAAGCCTCTCCCAAGCGCTT 681
173 rHisCysTrpIleSerGlyTrpIleSerIleGlnAspGlyValProLeuP 190
||||| ::::: ||||| |||
682 CCACTGACATGTCATGGGTGATGTGGCCCTCAGTGAAGCTCC 731

```

```

190 rHisProGlnThrLeuGlnLysLeuLysValProIleIleAspSerGln 206
||||| |||||::: ||||| |||
732 TGAGCCCAAGCCACTGCAGCAACTCGAGGTGCTGTGATCAGTCTGTAG 781
207 ValCysSerHisLeuTyrTrpArgGlyAla...GlyGlnGlyPro..... 220
||| ||||| |||
782 AGGTGATCTGCTGTACACATCGACGCCAGCCAGCTGAGAGCGGCACTT 831
221 .lIeThrGluAspMetLeuCysAlaGlyTyrLeuGlnGlyGluArgAspA 237
::: ||||| |||||::: ||||| |||
832 TGTCAGAGAGACATGTGTGTGCTGTGCTATGTGAGAGGGGCGCAAGGACG 881
237 lAcysLeuGlyAspSerGlyGlyProIleMetCysGlnValAspGlyAla 253
||||| ||||| ||||| |||
882 CCTCCAGGGTGACTCTGGGGGCCACTCTGCTGCTGTGAGAGGCTCTC 931
254 TrpLeuAlaGlyIleIleSerTrpGlyGlyGlyCysAlaGluArgAs 270
||| ||||| |||||::: ||||| |||
932 TGTAACCTGACGGGCAATTGTGAGCTGGGAGATGCCGTGTGGGCGCCGA 981
270 nArgProGlyValTyrIleSerLeuSerAlaHisArgSerTrpValGluL 287
||||| ||||| |||||::: ||||| |||
982 CAGGCTGTGTGTGTAACACTGTGGCTCCAGCTATGCTCTGATCCAA 1031
287 yslLeuAlaGlnGlyValGlnLeuArg 295
::: ||| ::::: |||
1032 GCAAGGTGACAGAACTCAGGCTGCT 1057

```











```

17 rphenrSerLeuLeuLeuAlaSerThrAlaIleuAsnAlaIa 34
   |||||||
77 CTTCACCTCCCTGCTGCTGCTGGCTGCACAGCATCTCATTCGGGCA 126
   |||||||
34 rglleProValProProAlaCysGlyLysProGlnGlnLeuAsnArgVal 50
   |||||||
127 GGATACCTGTTCCCGACAGCTGTGGAGAGCCCGACAGACTGACCGGGTT 176
   |||||||
51 ValGlyGlyGlyAspSerThrAspSerGluTrpProTrpIleValSerI 67
   |||||||
177 GTGGGGGGGAGAGACAGCAGCAGAGTGGCCCTGAGTCGTGACCAT 226
   |||||||
67 eGlnLysAsnGlyThrHisCysAlaGlySerLeuLeuThrSerArgT 84
   |||||||
227 CCAGAAAGATGGGACCCACACATGGGAGGTTCTGCTACACGCGGCT 276
   |||||||
84 rPValIleThrAlaAlaHisCysPheLysAspAsnLeuAsnLysProTyr 100
   |||||||
277 GGGTGAATCACTGCTGCCACTGTTTCAAGGACACCTGACAAACCAATAC 326
   |||||||
101 LeuPheSerValLeuLeuAlaTrpGlnLeuGlyAsnProGlySerAr 117
   |||||||
327 CTGTTCTGTGCTGCTGGGGGCTGGCAAGCTGGGGAAACCTGGCTCG 376
   |||||||
117 gSerGlnLysValGlyValAlaTrpValGluProHisProValLysSerT 134
   |||||||
377 GTCCGAAAGGTGGGTGTGCTGGGTGGAGCCACCCCTGTGTATCT 426
   |||||||
134 rPlyGlnGlnLysAlaAspIleAlaLeuValArgLeuGlnArgSer 150
   |||||||
427 GGAAGAGAGGCTGCTGCAGACATGGCTGCTGCTGCAGGCGCTCC 476
   |||||||
151 IlleGlnPheSerGluArgValLeuProIleCysLeuProAspAlaSerI 167
   |||||||
477 ATACAGTTCAGAGGGGCTCCGCCCATCTGCTACCTGATGCTTAT 526
   |||||||
167 eHisLeuProProAsnThrHisCysTrpIleSerGlyTrpGlySerIleG 184
   |||||||
527 CCACCTCCCTCCAAACACCCACTGAGATCTCAGGCTGGGGAGCATCC 576
   |||||||
184 LnsPglValPProleuProHisProGlnThrLeuGlnLysLeuLysVal 200
   |||||||
577 AAGTGGAGTTCCTTGGCCCAACCTCAGACCTGAGAGAGTGAAGGT 626
   |||||||
201 ProIleLeuAspSerGlnValCysSerHisLeuYrTrpArgIleAlaG 217
   |||||||
627 CCTATCATGACCTCGAAGTCTCAGCCCATCTGTACTGGCGGGAGCAGG 676
   |||||||
217 yGlnGlyProIleThrGlnAspIleLeuGlyLysArgIleGlnGlyG 234
   |||||||
677 ACAAGGACCATCATCAGAGCATGCTGTGTGGGCTACTTGGAGGGGG 726
   |||||||
234 LnsArgAspAlaCysLeuGlyAspSerGlyLysProLeuMetCysGlnVal 250
   |||||||
727 AGCGGATGCTGTGTGTGGGAGCTCCGGGGGCCCCCTCATGTGCCAGGTG 776
   |||||||
251 AspGlyAlaTrpLeuLeuAlaGlyIleIleSerTrpGlyGlnGlyVal 267
   |||||||
777 GACGGCGCTGGCTGTGCTGGCGGATCATCAGCTGGGGGAGGGGCTGC 826
   |||||||
267 agluArgAsnArgProGlyValTyrIleSerLeuSerAlaHisArgSerT 284
   |||||||
827 CGAGCGAACAAGGCCCGGGGTCTACATCAGACCTCTCTGCCACACGGTCT 876
   |||||||
284 rPValGlnLysIleValGlnGlyValGlnLeuArgGlyArgAlaGlnGly 300
   |||||||
877 GGGTGGAGAAAGATGCTGCAAGGGGTGAGCTCCGCGGGCGCTCAGGGG 926
   |||||||
301 GlyGlyAlaLeuArgAlaProSerGlnGlySerGlyAlaAlaIaArgSe 317
   |||||||
927 GGTGGGGCTTCAGGGGACCGAGGCAAGGCTGTGGGGCGCCCGCGGCTC 976

```

```

317 r 317
   |
977 C 977
seq_name: /SIDSI/gcgdata/geneseq/geneseqn.embl/NA2001A.DAT.AAS26942
seq_documentation_block:
ID AAS26942 standard; cdna; 1352 BP.
XX
AC AAS26942;
XX
DE 07-NOV-2001 (first entry)
XX
Human cDNA encoding a novel secreted protein, SEQ ID 134.
XX
Human; immunosuppressive; antiarthritic; ss; antirheumatic;
XX cytosolic; cardiant; vasotropic; cerebroprotective; nootropic;
XX neuroprotective; antibacterial; virucide; fungicide; ophthalmological;
XX vulnereary; secreted protein; rheumatoid arthritis;
XX hyperproliferative disorder; cardiovascular disorder; cardiac arrest;
XX cerebrovascular disorder; cerebral ischaemia; angiogenesis;
XX nervous system disorder; Alzheimer's disease; infection; ocular disorder;
XX corneal infection; wound healing; epithelial cell proliferation;
XX skin ageing; food additive; preservative; antiproliferative.
OS Homo sapiens.
XX
PN WO200155441-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01320.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-MAR-2000; 2000US-0198123.
PR 19-MAR-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.

```

1 MetValValSerGlyAlaProProAlaLeuGlyGlyGlyCysLeuGlyTh 17

```

|||||
27 ATGGTGGTTCTGGAGGGCCCCAGCCCTGGGTGGGGGCTGTCGGCAC 76
17 rPheThrSerLeuLeuLeuAlaSerThrAlaIleLeuAlaAla 34
77 CTTACACCTCCCTGCTGCTGGGTGGGTGACAGGCACTCTCAAGCGGCCA 126
34 rGlieProValProProAlaCysGlyLysProGlnGlnLeuAsnArgVal 50
127 GGATTAACCTGTTCCCTCCAGCTGTGGAGAGCCCAAGCAAGTGAACCGGTT 176
51 ValGlyGlyLeuAspSerThrAspSerGluTrpProThrIleValSer11 67
177 GTGGGGCGCGAGAGACAGCACTGACAGCAGAGTGGCCCTGATCGTAGCAT 226
67 eGlnLysAsnGlyThrHisCysAlaGlySerLeuLeuThrSerArg 84
227 CCAGAGAAGATGGAGCCCACTGCGAGGTTCTCTCTCAAGCCGCT 276
84 rPValIleThrAlaAlaHisCysPheLysAspAsnLeuAsnLysProTyr 100
277 GGGTATACACTGCTGCCACCTGTTCAAGGACAACTGAACAACCATAC 326
101 LeuPheSerValLeuLeuGlyAlaTrpGlnLeuGlyAsnProGlySerAr 117
327 CTGTTCTCTGCTGCTGGGGGCTGGGAGCTGGGGAACCTGGCTCTCG 376
117 gSerGlnLysValGlyValAlaTrpValGluProHisProValTyrSer 134
377 GTCCAGAGAGGTGGGTGTTGCTGGGTGGAGCCCAACCTGTATATCT 426
134 rPlyGlnGlyAlaCysAlaAspIleAlaLeuValArgLeuGluArgSer 150
427 GGAAGAGAGTGTCTGTGACAGCATTTGCCCTGTGCTGACGCTCC 476
151 IleGlnPheSerGluArgValLeuProIleCysLeuProAspAlaSer11 167
477 ATACAGTTCTCAGAGCGGTCTGCCACTGCTGCTGATGCTCTAT 526
167 eHisLeuProAsnThrHisCysTrpIleSerGlyTrpGlySer11leg 184
527 CCACCTCCCTCCAAACACCACTGTGATCTCAGGCTGGGGGAGCATCC 576
184 lAspGlyValProLeuProHisProGlnThrLeuGlnLysLeuVal 200
577 AAGATGAGTTCCTGCTGCCACCTCAGACCTCAGAACCTGAAGGTT 626
201 ProIleIleAspSerGluValCysSerHisLeuTyrTrpArgGlyAlaG1 217
627 CCTATATGACTCGAACTCTGCAACCATCTGTACTGCGGGAGGACAG 676
217 yGlnGlyProIleThrGluAspMetLeuCysAlaGlyTrpLeuGlnGly 234
677 ACAGGGAGCCATCATGAGACATGCTGTGCGGCTACTGTGAGGGGG 726
234 lArgAspAlaCysLeuGlyAspSerGlyGlyProLeuMetCysGlnAla 250
727 AGGGGATGTTCTGTGGGCGACTCCGGGGGCCCTCATGTGCGCAGGTT 776
251 AspGlyAlaTrpLeuLeuAlaGlyIleIleSerTrpGlyGlnGlyCysAl 267
777 GAGGGCCCTCGCTGTGCGCGGCATCATCAGCTGGGGGAGAGGCTGTGC 826
267 aGluArgAsnArgProGlyValTyrTrpLeuSerLeuSerAlaHisArgSer 284
827 CGAGCGCAACAGGCCCGGGGTCTACATCAGCCTCTCTGCGCAGCTCTCT 876
284 rPValGlnLysIleValGlnGlyValGlnLeuArgGlyArgAlaGlnGly 300
877 GGGTGAAGAAATCTGTGCAAGGGGTCACTCCGGGGGCGGCTCAGGGG 926
301 GlyGlyAlaLeuArgAlaProSerGlnGlySerGlyAlaAlaAlaArgSe 317
|||||

```

```

927 GGTGGGCGCTCAGAGGACCGAGCAGAGGCTGTGGGGCGCGCGGCTC 976
317 r 317
977 C 977

```

seq\_name: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:AAx87259

seq\_documentation\_block:  
ID AAX87259 standard; cDNA; 1378 BP.

AC AAX87259;

DT 27-SEP-1999 (first entry)

DE cDNA clone encoding human PRO343, amplified in tumour cells.

KW PRO343; UNQ302; cancer; tumour; diagnosis; therapy; human; ss.

OS Homo sapiens.

PH Key Location/Qualifiers

FT CDS 53..1006

FT sig\_peptide /\*tag- a

FT mat\_peptide /\*tag- b

FT /\*tag- c

PD WO9935170-A2.

PE 15-JUL-1999.

PF 05-JAN-1999; 99WO-US00106.

PR 20-NOV-1998; 98US-0109304.

PR 05-JAN-1998; 98US-0070440.

PR 29-APR-1998; 98US-0083500.

PR 22-MAY-1998; 98US-0086414.

PR 10-JUN-1998; 98US-0088742.

PR 10-NOV-1998; 98US-0107783.

PA (GENTH ) GENENTECH INC.

PI Botstein D, Goddard A, Gurney AL, Hillan KJ, Lawrence DA;

PI Roy MA, Wood WT;

DR WPI: 1999-430385/36.

DR P-PSDB; AAY06482.

PT Antibody against proteins expressed in neoplastic cells, useful for

PT tumor diagnosis and treatment

PS Example 1; Fig 11; 162pp; English.

PS This is the nucleotide sequence of cDNA clone DNA43318 (ATCC 209481)

CC coding for human PRO343 (UNQ302) (see AAY06482). The clone was

CC isolated from a foetal lung library. Amplification of DNA43318

CC (chromosome 16) was observed in primary lung and primary colon

CC tumours, suggesting an association with tumour formation or growth.

CC Antagonists (e.g. antibodies) directed against PRO343 may have

CC utility in cancer therapy. The invention identifies 14 genes (see

CC AAX87254-67) that are amplified in the genome of tumour cells. Such

CC amplification is expected to be associated with overexpression of

CC the gene product and to contribute to tumorigenesis. The encoded

CC proteins (see AAY06477-90) may be useful targets for the diagnosis

CC and/or treatment (including prevention) of certain cancers, and may

CC act as predictors of the prognosis of tumour treatment.

SQ Sequence 1378 BP; 235 A; 461 C; 412 G; 270 T; 0 other;

alignment\_scores:

seq_name: /SIDS1/gcgcdata/geneseq/geneseqn_emb1/NA1999.DAT.AAX52262	
seq_documentation_block:	
ID AAX52262 standard; DNA; 1378 BP.	
AC AAX52262;	
XX	
DT 25-JUN-1999 (first entry)	
XX	
DE Protein PR0343 CDNA clone DNA43318-1217.	
XX	
Secreted protein; transmembrane protein; human; enterocolitis;	
KN Zollinger-Ellison syndrome; gastrointestinal ulceration;	
KN congenital microvillus atrophy; skin disease; cell growth;	
KN abnormal keratinocyte differentiation; psoriasis; epithelial cancer	
KN Parkinson's disease; Alzheimer's disease; ALS; neuropathy;	
KN fibromodulin; dermal healing; Usher Syndrome; Atrophia areata;	
anti-thrombotic; wound healing; tissue repair; ss.	
XX	
OS Homo sapiens.	
XX	
PN W09914328-A2.	
XX	
PD 25-MAR-1999.	
XX	
PF 16-SEP-1998; 98MO-US19330.	
XX	
25-NOV-1997; 97US-0066840.	
PR 17-SEP-1997; 97US-0059113.	
PR 17-SEP-1997; 97US-0059115.	
PR 17-SEP-1997; 97US-0059117.	
PR 17-SEP-1997; 97US-0059119.	
PR 17-SEP-1997; 97US-0059121.	
PR 17-SEP-1997; 97US-0059122.	
PR 17-SEP-1997; 97US-0059184.	
PR 18-SEP-1997; 97US-0059263.	
PR 18-SEP-1997; 97US-0059266.	
PR 15-OCT-1997; 97US-0062125.	
PR 17-OCT-1997; 97US-0062285.	
PR 21-OCT-1997; 97US-0062287.	
PR 24-OCT-1997; 97US-0062814.	
PR 24-OCT-1997; 97US-0062816.	
PR 24-OCT-1997; 97US-0063045.	
PR 24-OCT-1997; 97US-0063120.	
PR 24-OCT-1997; 97US-0063121.	
PR 24-OCT-1997; 97US-0063127.	
PR 24-OCT-1997; 97US-0063128.	
PR 27-OCT-1997; 97US-0063329.	
PR 27-OCT-1997; 97US-0063327.	
PR 28-OCT-1997; 97US-0063541.	
PR 28-OCT-1997; 97US-0063542.	
PR 28-OCT-1997; 97US-0063544.	
PR 28-OCT-1997; 97US-0063549.	
PR 28-OCT-1997; 97US-0063550.	
PR 28-OCT-1997; 97US-0063564.	
PR 29-OCT-1997; 97US-0063565.	
PR 29-OCT-1997; 97US-0063575.	
267 agluAgaSuarPrGlyValTyrlSerLeuSerAlaHisArgSer 284	
853 CGAGCCCAACAGCGCCGGGGCTTACATACAGCTCTCTGCGACCGCTCT 902	
284 rPVaLgLuLysIleValGlnGlyValGlnLeuArgGlyArgGlaGlnGly 300	
903 GGGTGAAGAAGATCGTGCAGGGGGTGCAGCTCCCGGGCGGGCTCAGGG 952	
301 GLGfLAlaLeuArgAlaProSerGlnGlySerGlyValAlaAlaArgse 317	
953 GGGGGGGCCCTCAGGGGCACCGACCGAGGCTCTGGGGCCGCGGCGCTC 1002	
317 r 317	
1	
1003 C 1003	

PR 29-OCT-1997; 97US-0063704.  
 PR 29-OCT-1997; 97US-0063732.  
 PR 29-OCT-1997; 97US-0063734.  
 PR 29-OCT-1997; 97US-0063738.  
 PR 29-OCT-1997; 97US-0064215.  
 PR 29-OCT-1997; 97US-0063735.  
 PR 31-OCT-1997; 97US-0063870.  
 PR 31-OCT-1997; 97US-0064103.  
 PR 03-NOV-1997; 97US-0064248.  
 PR 07-NOV-1997; 97US-0064809.  
 PR 12-NOV-1997; 97US-0065186.  
 PR 17-NOV-1997; 97US-0065846.  
 PR 18-NOV-1997; 97US-0065693.  
 PR 21-NOV-1997; 97US-0066120.  
 PR 21-NOV-1997; 97US-0066364.  
 PR 24-NOV-1997; 97US-0066772.  
 PR 24-NOV-1997; 97US-0066466.  
 PR 24-NOV-1997; 97US-0066770.  
 PR 24-NOV-1997; 97US-0066511.  
 PR 24-NOV-1997; 97US-0066453.

(GERTH ) GENENTECH INC.

XX Chen J, Goddard A, Gurney AL, Pennica D, Wood WI, Yuan J;  
 FI WPI; 1999-229533/19.  
 DR P-PSDB; AAY13391.  
 DR

PT New isolated human genes and polypeptides used in, e.g. treatment of  
 PT gastrointestinal ulceration

XX Claim 2; Fig 97; 320p; English.

XX AAX52213-74 encode secreted and transmembrane fetal proteins, and are  
 CC obtained from cDNA libraries, prepared from fetal lung, fetal kidney,  
 CC fetal brain, fetal liver and fetal retina. The encoded polypeptides  
 CC have specific uses based on their homology to known polypeptides.  
 CC e.g. PRO211 and PRO217 can be used for disorders associated with the  
 CC preservation and maintenance of gastrointestinal mucosa and the repair  
 CC of acute and chronic mucosal lesions (e.g. enterocolitis,  
 CC Zollinger-Ellison syndrome, gastrointestinal ulceration and congenital  
 CC microvillus atrophy), skin diseases associated with abnormal  
 CC keratinocyte differentiation (e.g. psoriasis, epithelial cancers such as  
 CC lung squamous cell carcinoma of the vulva and gliomas), potent effects on  
 CC cell growth and development, diseases related to growth or survival of  
 CC nerve cells including Parkinson's disease, Alzheimer's disease, ALS,  
 CC neuropathies or cancer. PRO265 can be used as for fibromodulin, e.g. for  
 CC reducing dermal scarring. PRO264 can be used as a target for anti-tumor  
 CC drugs. PRO533 may be used in the treatment of Usher Syndrome or Atrophila  
 CC areata. PRO269 can be used as an anti-thrombotic agent. PRO287  
 CC polypeptides and portions may have therapeutic applications in wound  
 CC healing and tissue repair. PRO317 can be used for treating problems of  
 CC the kidney, uterus, endometrium, blood vessels, or related tissue, e.g.  
 CC in the heart of genital tract.

XX Sequence 1378 BP; 235 A; 461 C; 412 G; 270 T; 0 other;

alignment\_scores:  
 Quality: 1720.00 Length: 317  
 Ratio: 5.426 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-10-040-803-7 x AAX52262 ..

Align seg 1/1 to: AAX52262 from: 1 to: 1378

1 MetValValSerGlyAlaProProAlaLeuGlyGlyGlyCysLeuGlyYth 17  
 |||||  
 53 ATGGTGGTTTCTGAGCGCCCGCCGCTGGGCTGTCTCGGCAC 102  
 |||||  
 17 rphrSerLeuLeuLeuAlaSerThrAlaIleLeuAsnAlaAla 34

|||||  
 103 CTTACCTCCCTGCTGCTGGGCTGCAGACCATCTCAATGCGGCCA 152  
 34 rglProValProProAlaCysGlyYsProGlnGlnLeuAsnAlaVal 50  
 |||||  
 153 GGATACCTGTTCCCGCCAGCCTGTGGAAAGCCCGAGAGCTGAACCGGGTT 202  
 51 ValGlyGlyCysAspSerThrAspSerGluThrProThrIleAlaSer1 67  
 203 GTGGCGCGCGAGACAGCAGTGCAGAGGAGGAGGCTGCTGCTGAGCAT 252  
 67 eGlnLysAsnGlyThrHisHisCysAlaGlySerLeuLeuThrSerArg 84  
 253 CCAAGAAATGAGACCCACCACTGCGAGGTTCTCTCCACAGCGCGCT 302  
 84 rValIleThrAlaAlaHisCysPheLysAspAsnLeuAsnLysProTy 100  
 303 GGTGATCACTGCTGCCACATGTTCAAGGACAACCTGAACCAACATAC 352  
 101 LeuPheSerValLeuLeuGlyAlaThrGlnLeuGlyAsnProGlySer 117  
 353 CTGTTCTGCTGCTGCTGGGCGCTGCGAGCTGGGGAACCTGGCTGTG 402  
 117 gSerGlnLysValGlyValAlaThrValGluProHisProValTyrSer 134  
 403 GTCCAGAAAGTGGGTGTTCCTGGGTGAGCCCGCCCTGTGATTCCT 452  
 134 rPlyGlnGlyAlaCysAlaAspIleAlaLeuValAlaGlnGluArgSer 150  
 453 GGAAGAAAGTGGCTGTGCAACATGCTGCTGGTGGTCTGAGAGCTTC 502  
 151 IleGlnPheSerGluArgValLeuProIleCysLeuProAspAlaSer 167  
 503 ATACAGTTTCAGAGCGGCTCTGCCATCTGCCTACCTGATGCTCTAT 552  
 167 eHisLeuProProAsnThrHisCysThrIleSerGlyTyrGlySerIle 184  
 553 CCACCTCCCTCAACACCCACTGCTGATCTCAGCGTGGGAGCATCC 602  
 184 lAspGlyValProLeuProHisProGlnThrLeuGlnLysVal 200  
 603 AAGATGAGTTCCTTGGCCACCTCAGACCTCAGACCTCAGACCTAAGTT 652  
 201 ProIleAspSerGluValCysSerHisLeuTyrTrpArgIleAla 217  
 653 CCTATCATGACTCGGAAGTGTGAGCCATCTGTGCGGGAGCAGG 702  
 217 yGlnGlyProIleThrGluAspMetLeuCysAlaGlyTyrLeuGlnGly 234  
 703 ACAGGAGCCATCAGAGGACATCTGTGCGGCTACTGAGGGGG 752  
 234 lAspAspAlaCysLeuGlyAspSerGlyGlyProLeuMetCysGlnVal 250  
 753 AGCGGATGCTGTCTGGGCGACTCCGGGGGCCCTCATGTGCAAGGTG 802  
 251 AspGlyAlaThrLeuLeuAlaGlyIleIleSerTrpGlyGlnGlyCysAl 267  
 803 GACGGCGCTGCTGCTGCGCGCATCATCAGCTGGCGAGGCTGTGC 852  
 267 aglUArgAsnArgProGlyValTyrTrleSerLeuSerAlaHisArgSer 284  
 853 CGAGCCACAGGCGCGGGGTCTACTACGCTCTCTCGCAGCCGCTCT 902  
 903 GGGTGGAGAAAGATCGTCAAGGGGTGACACTCCGGGGCGCGCTAGGGG 952  
 284 rValGlnLysIleValGlnGlyValGlnLeuArgGlyArgAlaGlnGly 300  
 301 GlyGlyAlaLeuArgAlaProSerGlnGlySerGlyAlaAlaAlaArg 317  
 953 GTTGGGGCTCTCAGGGACCGAGCGAGGCTGTGGGGCGCGCGCTC 1002  
 317 r 317

1003 C 1003

seq\_name: /SIDSL/gcgdata/geneseq/geneseq\_emb1/NA2000.DAT:AAA46914

seq\_documentation\_block:

ID AAA46914 standard; cDNA; 1378 BP.

XX AAA46914;

DT 03-OCT-2000 (first entry)

DE cDNA encoding novel polypeptide PRO343.

KW PRO201; PRO292; PRO327; PRO1265; PRO344; PRO343; PRO347; PRO357;

KW PRO715; PRO1017; PRO1112; PRO509; PRO853; PRO882; tumour cell;

KW tumorigenesis; cancer; neoplastic cell growth; cell proliferation; ss.

XX Homo sapiens.

FH Key Location/Qualifiers

CDS 53..1007

/\*tag= a

WO200037640-A2.

XX 29-JUN-2000.

PF 16-DEC-1999; 99WO-US30095.

PR 22-DEC-1998; 98US-0113296.

PR 08-MAR-1999; 99WO-US05028.

PR 02-JUN-1999; 99WO-US12252.

PR 01-SEP-1999; 99WO-US20111.

PR 15-SEP-1999; 99WO-US21090.

PR 30-NOV-1999; 99WO-US28313.

PR 30-NOV-1999; 99WO-US28409.

PR 01-DEC-1999; 99WO-US28301.

PR 02-DEC-1999; 99WO-US28565.

XX (GETH ) GENENTECH INC.

PI Botstein D, Goddard A, Gurney AL, Hillan K, Lawrence DA, Roy MA;

XX Wood WT;

DR WPI: 2000-452188/39.

DR P-PSDB; AAY93689.

XX New anti-polypeptide antibody useful in the treatment and diagnosis of

neoplastic cell growth and proliferation -

Claim 50; Fig 11; 220p; English.

XX The present sequence encodes a novel human polypeptide. The

CC specification describes novel polypeptides designated PRO201, PRO292,

CC PRO327, PRO1265, PRO344, PRO343, PRO347, PRO357, PRO1017,

CC PRO1112, PRO509, PRO853 and PRO882. These genes are amplified in

CC the genome of tumour cells. The polypeptides are believed to contribute

CC to tumorigenesis. The polypeptides are useful target for the

CC identification of certain cancers, and may act as predictors of the

CC prognosis of tumour treatment. Antibodies against these polypeptides

CC are useful in the treatment and diagnosis of neoplastic cell growth

CC and proliferation in mammals.

XX Sequence 1378 BP; 235 A; 461 C; 412 G; 270 T; 0 other;

XX

alignment\_scores:

Quality: 1720.00

Ratio: 5.426

Percent Similarity: 100.000

Length: 317

Gaps: 0

Percent Identity: 100.000

alignment\_block:

US-10-040-803-7 x AAA46914

Align seg 1/1 to: AAA46914 from: 1 to: 1378

```
1 MetValValSerGlyAlaProProAlaLeuGlyGlyCysLeuGlyTh 17
|||||
53 ATGGTGTTTCTTGAGAGCGCCCGACCGCTGGGTGGGGCTGTCTCGCAC 102
|||||
17 rPheThrSerLeuLeuLeuLeuAlaSerThrAlaIleLeuSnaIaIa 34
|||||
103 CTTACCTCTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 152
|||||
34 rGllLeProValProProAlaCysGlyLysProGlnGlnLeuSnaIaIa 50
|||||
153 GGATACCTCTTCCCGACCGCTGGGAAAGCCCGACAGCTGAGCGGGTT 202
|||||
51 ValGlyGlyGlyLysSerThrAspSerGluTrpProTyrIleValSer 67
|||||
203 GTGGCGCGGAGAGACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 252
|||||
67 eGlnLysAsnGlyThrHisCysAlaGlySerLeuLeuThrArgT 84
|||||
253 CCAGAGAGATGGGACCCACACCTGCGAGGTTCTCTCTCACACGCGCT 302
|||||
84 rPValIleThrAlaAlaHisCysPheLysAspAsnLeuAsnLysProTyr 100
|||||
303 GGGTATCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 352
|||||
101 LeuPheSerValLeuLeuGlyAlaTrpGlnLeuGlyAsnProGlySerAr 117
|||||
353 CTGTTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 402
|||||
117 gSerGlnLysValGlyAlaIleTrpValGluProHisProValLysSerT 134
|||||
403 GTCCAGAAAGTGGGTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 452
|||||
134 rPlyGlyGlyAlaCysAlaAspIleAlaLeuValArgLeuGlyLysSer 150
|||||
453 GGAAGGAAGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 502
|||||
151 IleGlnPheSerGluArgValLeuProIleCysLeuProAspAlaSerI 167
|||||
503 ATACAGTTTCTCAGAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 552
|||||
167 eHisLeuProProAsnThrHisCysTrpIleSerGlyTrpGlySerIleG 184
|||||
553 CCACCTCCCTCAAAACACCACTGCTGATCTCAGGCTGGGAGCATCC 602
|||||
184 LAspGlyValProLeuProHisProGlnThrLeuGlnLysVal 200
|||||
603 AAGATGAGATTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 652
|||||
201 ProIleLeuAspSerGluValCysSerHisLeuTrpTrpArgIleVal 217
|||||
653 CTTATCATCTCAGAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 702
|||||
217 yGlnGlyProIleThrGlnAspMetLeuCysAlaGlyTyrLeuGlnGly 234
|||||
703 ACAGAGACCATCATCTAGAGCATGCTGCTGCTGCTGCTGCTGCTGCTGCT 752
|||||
234 LArgAspAlaCysLeuGlyAspSerGlyLysProLeuMetCysGlnVal 250
|||||
753 AGCGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 802
|||||
251 AspGlyAlaTrpLeuLeuAlaGlyIleIleSerTrpGlyGlyCysAl 267
|||||
803 GACGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 852
|||||
267 aGluArgAsnArgProGlyValTyrIleSerLeuSerAlaHisArgSer 284
|||||
853 CGAGCGGAACAGGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 902
|||||
284 rPValGlyLysIleValGlnGlyValGlnLeuArgGlyArgAlaGlnGly 300
|||||
```

```

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAF72420
seq_documentation_block:
ID      AAF72420 standard; cDNA; 1378 BP.
XX      AAF72420;
AC
DT      24-APR-2001 (first entry)
Human PRO343 cDNA.
Human, PRO; dermatological; antipsoriatic; cytostratic; antiinflammatory;
antiparkinsonian nootropic; neuroprotective; vulnerrary; radiant;
antiangiogenic; vasotrophic; antiasthmatic; antirheumatic; cancer;
Kw      antiatheritic; antifertility; antidiabetic; antiviral; diabetes;
KW      ophthalmological; gene therapy; skin disease; gastrointestinal disorder
ischaemia; inflammation; ss.
XX
OS      Homo sapiens.
XX      WO200104311-A1.
PN
PD      18-JAN-2001.
PE      22-FEB-2000; 2000WO-US04414.
XX
PR      07-JUL-1999; 99US-0143048.
PR      26-JUL-1999; 99US-0145638.
PR      28-JUL-1999; 99US-0146222.
PR      08-SEP-1999; 99WO-US20594.
PR      13-SEP-1999; 99WO-US20944.
PR      15-SEP-1999; 99WO-US21090.
PR      15-SEP-1999; 99WO-US21547.
PR      05-OCT-1999; 99WO-US23089.
PR      29-NOV-1999; 99WO-US28214.
PR      30-NOV-1999; 99WO-US28313.
PR      16-DEC-1999; 99WO-US30095.
PR      20-DEC-1999; 99WO-US30911.
PR      20-DEC-1999; 99WO-US30999.
PR      05-JAN-2000; 99WO-US00219.
XX
PA      (GETH ) GENENTECH INC.
XX
PI      Ashkenazi AJ, Botstein D, Desnoyers L, Eaton DL, Ferrara N;
PI      Fildoroff E, Fong S, Gao W, Gerber H, Gerritsen ME, Goddard A;
PI      Goldowski PJ, Grimaldi CJ, Gurney AL, Hillan KJ, Kljavin IJ;
PI      Mather JP, Pan J, Paoni NF, Roy MA, Stewart TA, Tumas D;
PI      Williams PM, Wood WI;
XX
WI      MPI, 2001-081051/09.
DR      P-PSDB; AAB80259.
XX
PT      Sixty one nucleic acids encoding PRO polypeptides which are useful in
PT      the treatment of skin diseases (e.g. psoriasis), cancers (e.g. lung
PT      squamous cell carcinoma) and neurodegenerative diseases (e.g.
PT      Alzheimer's disease) -
XX
PS      Claim 2; Fig 97; 393pp; English.
CC      The present sequence is one of sixty one nucleic acids encoding novel
CC      secreted and transmembrane PRO polypeptides. The PRO polypeptides are
CC      useful for treating skin diseases (e.g. psoriasis), cancers (e.g. lung
CC      squamous cell carcinoma), gastrointestinal disorders (e.g.

```

CC enterocolitis), neurodegenerative diseases (e.g. Alzheimer's disease,  
CC Parkinson's disease), wound repair, cardiovascular disorders (e.g.  
CC endothelial bleeding angiogenesis, ischemias such as coronary  
CC ischemia, atherosclerosis), inflammatory disorders (e.g. asthma,  
CC rheumatoid arthritis, multiple sclerosis), infertility, AIDS and  
CC diabetes and retinal disorders such as retinitis pigmentosa.  
CC The PRO nucleic acids have applications in molecular biology, including  
XX use as hybridization probes, and in chromosome and gene mapping.

SQ Sequence 1378 BP; 235 A; 461 C; 412 G; 270 T; 0 other;

Alignment\_scores:

	Quality: 1720.00	Length: 317
	Ratio: 5.426	Gaps: 0
Percent Similarity: 100.000	Percent Identity: 100.000	

alignment\_block:  
US-10-040-803-7 x AAF72420 ..

Align seg 1/1 to: AAF72420 from: 1 to: 1378

```
1 MetValAlSerGIyAlaProProAlaLeuGlyGlyGlyCysLeuGlyTh 17
|||||
53 ATGGTGGTTTCTGGAGCGGCCCAAGCCCTGTGGGGGCTGACTGGGAC 102
|||
17 rPheThrSerLeuLeuLeuAlaSerThrAlaIleLeuAsnAlaAla 34
|||||
103 CTTCACCTCCCTGCCTGCTGGCGCTGCAGACGACATCTCAATGGGACA 152
|||||
34 rGILProValProProAlaCysGLyLysProGlnGlnLeuAsnArGyl 50
|||||
153 GGATTACCTGTTCCCGCACGCTGTGGGAAGGCCACAGACGCTGAACGGGGTT 202
|||||
51 ValGlGlyGlyLuAspSerThrAspSerGluTrpProTrpIleValSer1 67
|||||
203 GTGGGGGCGAGAGACAGCATGACAGGAGTGAGTGGCCCTGTGATGCTGAGCAT 252
|||||
67 eGlnLysAsnGlyThrHisHIScysAlaGlySerLeuLeuThrSerArgT 84
|||||
253 CCAGAAGAAATGGGACCACACACTGCGAAGTTCTTCTGCTCACCAGCGCT 302
|||||
84 rPvalIleThrAlaIleHisCysPhelYaspsAsnuAsnuLysProTyR 100
|||||
303 GGGTGATCATCTGCCCCACGTGTTCAAGGACAACCTGAACAACCATATAC 352
|||||
101 LeuPheSerValLeuLeuGlyAlaIleTrpGlnLeuGlyAsnProGlySerAr 117
|||||
353 CTGTTCTCTGTGCTGCTGGGGGCTCGGAGGTGGGGAACCTTGACTCTCG 402
|||||
117 gSerGlnLysValGlyValAlaIleTrpValGluProHisProValTyrSerT 134
|||||
403 GTCCCAAGAAGTGGGTGTGGCTGGGTGGAGACCCACCTGTATATTCTT 452
|||||
134 rPlYSgluGlyAlaCysAlaSpIIleAlaLeuValAlaGluGluAsr 150
|||||
453 GGAAGAGAAGTGCCTGTGCAGACATTTGCCCTGGGCTTCGAGCGCTCC 502
|||||
151 IleGlnPheSerGluArgValLeuProIleCysLeuProAspAlaSerIl 167
|||||
503 ATTACAGTTCTCAGACGGGCTCTGTGCCATCTGCTACTGATGCTGCTATF 552
|||||
167 eHisLeuProProAsnThrHisCysTrpIleSerGlyTrpGlySerIleG 184
|||||
553 CCACACTGCCCTCCAACACCCATCTGTGATTCAGGCTGGGGAGAGATCC 602
|||||
184 LnsApGlyValProLeuProHisProGlnThrLeuGlnLysLeuVal 200
|||||
603 AAGTGGAGTTCCTTGCCCCACCTTCAGACCTTCGAGAAAGCTGAAGTT 652
|||||
201 ProIleIleAspSerGluValCysSerHisLeuTyrrTrpArgIyalagl 217
|||||
653 CCTATATATGACTCGGAAGTCTGCAGCATCTGTACTGCGGGGAGCAGG 702
```

```

217 yGlnGlyProIleThrGluAspMetLeuCysAlaGlyTyrLeuGly 234
|||||
703 ACAGGAGCCATCACTGAGGACATGCTGTGCGGCTACTTGAAGGGG 752
224 luATgAspAlaCysLeuGlyAspSerGlyGlyProLeuMetCysGlnVal 250
753 AGCGGAGATGTTGTCGGGCGACTCCGGGGCCCCCTCATGTGCCAGGTG 802
251 AspGlyAlaTrpLeuLeuAlaGlyTyrLeuSerTrpGlyGlyCysAl 267
803 GACGGCGCTGCTGCTGGCGGATCATCATGAGGGCGGCGGCTGTGC 852
267 agluARsnaRProGlyValTyrIleSerLeuSerAlaHisArgSer 284
853 CGAGGGCAACAGCGCCGGGCTCATCATGCTCTGCGGCGCTCTCT 902
284 rPVAlGluValIleValGlnGlyValGlnLeuArgGlyArgAlaGlnGly 300
903 GGGTGGAGAGATGCTGCAAGGGGTGCAAGCGCGGCGGCTCAAGGG 952
301 GlyGlyAlaLeuArgAlaProSerGlnGlySerGlyAlaAlaArgse 317
953 GGTGGGCGCCCTCAGGGGACCGAGCCAGGCTGTGGGGCGCGCGCTC 1002
317 r 317
1003 C 1003

seq_name: /STD1/seqdata/geneseq/geneseqn-emb1/NA2001A.DAT.AAD02990
seq_documentation_block:
ID AAD02990 standard; cDNA; 1430 BP.
XX
AC AAD02990;
XX
DT 31-MAY-2001 (first entry)
XX
DE Human serine protease, protease C-E cDNA.
XX
KW Human; serine protease; protease C-E; therapy; desquamation; skin care;
KW laundry detergent; shampoo; cleaning agent; hair care; skin flaking;
KW neurodegenerative disorder; dermatological; immunogenic; proteolytic;
KW chromosome 16p13.3; ss.
XX
OS Homo sapiens.
XX
XX
Key Location/Qualifiers
CDS 113..1066
/label= a
/feature= "Human serine protease, protease C-E"
XX
XX WO200116288-A2.
XX
XX 08-MAR-2001.
XX
XX 14-AUG-2000; 2000WO-US22117.
XX
XX 31-AUG-1999; 99US-0386629.
XX
XX (ORTH ) ORTHO-MCNEIL PHARM INC.
XX
XX Darrow A, Qi J, Andrade-Gordon P;
XX
XX WPI; 2001-226681/23.
XX
XX P-PSDB; AAY72890.
XX
XX Novel serine protease termed protease C-E, useful for treating and
XX preventing skin flaking or imbalance of desquamation -
XX
XX Claim 2; Fig 1; 78pp; English.
XX
XX The present sequence is a human serine protease, protease C-E cDNA which

```

CC is a member of the S1 serine protease family. Protease C-E gene is located on chromosome 16p13.3 and is expressed in pancreas, placenta, prostate, small intestine, stomach, spleen, fibroblasts, epidermis, cerebellum, cerebral cortex, pituitary and hippocampus. Protease C-E is useful for treating an imbalance of desquamation, by topical application. A non-pharmaceutical composition comprising the protein may be formulated as a laundry detergent, shampoo, hair care composition and hair care composition. Protease C-E is useful for treating and preventing skin flaking, neurodegenerative disorders and dermatological pathologies. It is less immunogenic to sensitive individuals and it provides efficient proteolytic activity in a non-natural environment.

Sequence 1430 BP; 240 A; 484 C; 428 G; 278 T; 0 other.

#### alignment\_scores:

Quality: 1720.00 Length: 317  
Ratio: 5.426 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

#### alignment\_block:

US-10-040-803-7 x AAD02990

Align seg 1/1 to: AAD02990 from: 1 to: 1430

```

1 MetValSerGlyAlaProProAlaLeuGlyGlyGlyCysLeuGlyTh 17
|||||
113 ATGGTGGTTCGTGGAGCGCGCCAGCCCTGGGTGGGGCTGTCTGGGAC 162
17 rPheThrSerLeuLeuLeuAlaSerThrAlaIleLeuAsnAlaAla 34
|||||
163 CTTCACCTCCCTGCTGCTGCTGGCTGCAGACCATCTCATATGGGGCA 212
34 rGleProValProProAlaCysGlyLysProGlnGlnLeuAsnArgVal 50
213 GGATACCTGTGTCCTCCAGCTGTGGAGAGCCCAACACACTGAACGGTT 262
51 ValGlyGlyGlyAspSerThrAspSerIuTrpProIleValSerI 67
263 GGGGCGCGGAGAGACAGACAGCAGCGAGCGCCCTGGATCGTAGCAT 312
67 eGlnLysAsnGlyThrHisCysAlaGlySerLeuLeuThrSerArg 84
313 CCAGAGAAATGGAGACCAACACACAGTGTCTCTGTCACCAAGCCCT 362
84 rPVAlIleThrAlaAlaHisCysPheLysAspAsnLeuAsnLysProTyr 100
363 GGGTATCACTGCTGCCCATGTTTCAAGGACAACTGAACAAACCATAC 412
101 LeuPheSerValLeuLeuGlyAlaTrpGlnLeuLysAsnProGlySerAr 117
413 CTGTTCTCTGTGCTGCTGGGGCTGGGAGCTGGGGAACCTGGCTCTCG 462
117 gSerGlnLysValGlyValAlaTrpValGluProHisProValTyrSer 134
463 GTCCAGAGAGGTGGGTGTTGCTGGGTGGAGCCCAACCTGTGATTCCT 512
134 rPVysGluGlyAlaCysAlaAspIleAlaLeuValArgLeuGluArgSer 150
513 GGAAGGAAGGTGCTGTGCAGACATTCCTGTGGCTGTGCAGAGCGCTCC 562
151 IleGlnPheSerGluArgValLeuProIleCysLeuProAspAlaSerI 167
563 ATACAGTTCTAGAGCGGGTCTGCTCCATCTGCTACTCTATGCTCTAT 612
167 eHisLeuProProAsnThrHisCysTrpIleSerGlyTyrGlySerIle 184
613 CCACCTCCCTCCAAACACCACTGCTGAGTCAAGGTGGGAGGATCC 662
184 lnaSPGlyValProLeuProHisProGlnThrLeuGlnLysLeuVal 200
663 AAGATGAGATTCCCTTCCACCTCAGACCTCGAGACCTGCAAGAGTGAAGTT 712

```



201 ProIleIleAspSerGIuValCysSerHisLeuTYrTrpArgGIuAlaGI 217  
|||||  
713 CCTATCATGACTCGAAGTCTGCAGCATCTGTACTGGCGGAGACAGG 762  
217 YGIuGIProIleThrGIuAspMetLeuCysAlaGIYrLeuGIuGIY 234  
|||||  
763 ACAGGACCATCATCTGAGACATGCTGTGGCGGCTACTTGGAGGGCG 812  
234 IuArgAspAlaCysLeuGIYAspSerGIYGIYProIleMetCysGIuVal 250  
|||||  
813 AGCGGATGCTGTGTCTGGGACTCCGGGGCCCCCTCATGTGCCAGGTG 862  
251 AspGIYAlaTrpLeuAlaGIYIleSerTrpGIYGIuGIYCysAl 267  
|||||  
863 GACGGCCCTGGCTGGCTGGCCGATCATGCTGGGGGAGGGCTGTGC 912  
267 AGIuArgAsnArgProGIYAlaTYrIleSerLeuSerAlaHisArgSer 284  
|||||  
13 CGACGGCAGACAGCGCCGGGCTACATCAGCCCTCTGGCGACGGCTCT 962  
284 rPValGIuYsIleValGIuGIYAlaGIuAlaArgGIYArgAlaGIuGIY 300  
|||||  
963 GGGTGGAGAGATGCTGCAAGGGGTGCAGCTCCGGGGCGCGCTCAGGGG 1012  
301 GIYGIYAlaLeuArgAlaProSerGIuGIYSerGIYAlaAlaArgSe 317  
|||||  
1013 GGTGGGGCCCTCAGGGCAGCCAGGCGCTGTGGGGCCCGCGCGCTC 1062  
317 r 317  
1063 C 1063

seq\_name: /STDS1/gcgdata/geneseq/geneseqn\_emb1/NA2001A.DAT.AAS26871  
seq\_documentation\_block:  
ID AAS26871 standard; cDNA; 1439 BP.

XX AAS26871:

DT 07-NOV-2001 (first entry)

DE Human cDNA encoding a novel secreted protein, SEQ ID 63.

XX Human; immunosuppressive; antiarthritis; ss: antineumatic;

cytostatic; cardiant; vasotropic; cerebroprotective; nootropic;

neuroprotective; antibacterial; virucide; fungicide; ophthalmological;

hyperproliferative disorder; rheumatoid arthritis;

hyperproliferative disorder; cardiovascular disorder; cardiac arrest;

cerebrovascular disorder; cerebral ischemia; angiogenesis;

nervous system disorder; Alzheimer's disease; infection; ocular disorder;

corneal infection; wound healing; epithelial cell proliferation;

skin ageing; food additive; preservative; antiproliferative.

XX Homo sapiens.

OS WO200155441-A2.

XX 02-AUG-2001.

XX 17-JAN-2001; 2001WO-US01320.

XX 31-JAN-2000; 2000US-0179065.

XX 04-FEB-2000; 2000US-0180628.

XX 24-FEB-2000; 2000US-0184664.

XX 02-MAR-2000; 2000US-0186350.

XX 16-MAR-2000; 2000US-0189874.

XX 17-MAR-2000; 2000US-0190076.

XX 18-APR-2000; 2000US-0198123.

XX 19-MAY-2000; 2000US-0205515.

XX 07-JUN-2000; 2000US-0209467.

XX 28-JUN-2000; 2000US-0214886.

XX 30-JUN-2000; 2000US-0215135.

PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217486.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225477.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226868.  
PR 22-AUG-2000; 2000US-0227162.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236366.  
PR 29-SEP-2000; 2000US-0236366.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239933.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.

PR 20-OCT-2000; 2000US-0241808.  
 PR 20-OCT-2000; 2000US-0241809.  
 PR 20-OCT-2000; 2000US-0241826.  
 PR 01-NOV-2000; 2000US-0244617.  
 PR 08-NOV-2000; 2000US-0246474.  
 PR 08-NOV-2000; 2000US-0246475.  
 PR 08-NOV-2000; 2000US-0246476.  
 PR 08-NOV-2000; 2000US-0246477.  
 PR 08-NOV-2000; 2000US-0246523.  
 PR 08-NOV-2000; 2000US-0246524.  
 PR 08-NOV-2000; 2000US-0246525.  
 PR 08-NOV-2000; 2000US-0246526.  
 PR 08-NOV-2000; 2000US-0246527.  
 PR 08-NOV-2000; 2000US-0246528.  
 PR 08-NOV-2000; 2000US-0246533.  
 PR 08-NOV-2000; 2000US-0246609.  
 PR 08-NOV-2000; 2000US-0246610.  
 PR 08-NOV-2000; 2000US-0246611.  
 PR 08-NOV-2000; 2000US-0246613.  
 PR 17-NOV-2000; 2000US-0249207.  
 PR 17-NOV-2000; 2000US-0249208.  
 PR 17-NOV-2000; 2000US-0249209.  
 PR 17-NOV-2000; 2000US-0249210.  
 PR 17-NOV-2000; 2000US-0249211.  
 PR 17-NOV-2000; 2000US-0249212.  
 PR 17-NOV-2000; 2000US-0249213.  
 PR 17-NOV-2000; 2000US-0249214.  
 PR 17-NOV-2000; 2000US-0249215.  
 PR 17-NOV-2000; 2000US-0249216.  
 PR 17-NOV-2000; 2000US-0249217.  
 PR 17-NOV-2000; 2000US-0249218.  
 PR 17-NOV-2000; 2000US-0249244.  
 PR 17-NOV-2000; 2000US-0249245.  
 PR 17-NOV-2000; 2000US-0249264.  
 PR 17-NOV-2000; 2000US-0249265.  
 PR 17-NOV-2000; 2000US-0249297.  
 PR 17-NOV-2000; 2000US-0249299.  
 PR 17-NOV-2000; 2000US-0249300.  
 PR 01-DEC-2000; 2000US-0250160.  
 PR 01-DEC-2000; 2000US-0250391.  
 PR 05-DEC-2000; 2000US-0251030.  
 PR 05-DEC-2000; 2000US-0251988.  
 PR 05-DEC-2000; 2000US-0256719.  
 PR 06-DEC-2000; 2000US-0251479.  
 PR 08-DEC-2000; 2000US-0251856.  
 PR 08-DEC-2000; 2000US-0251868.  
 PR 08-DEC-2000; 2000US-0251869.  
 PR 08-DEC-2000; 2000US-0251990.  
 PR 11-DEC-2000; 2000US-0254097.  
 PR 05-JAN-2001; 2001US-0259678.  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 PI Rosen CA, Barash SC, Ruben SM;  
 XX  
 DR WPI; 2001-476222/51.  
 XX P-PSDB; AAU16966.  
 XX  
 PT Novel polypeptides and polynucleotides useful as diagnostic reagents to  
 PT diagnose diseases or disorders associated with aberrant expression or  
 PT activity of polypeptides, for treating blood clotting disorder,  
 PT haemophilia  
 XX  
 PS Claim 1; SEQ ID No 63; 601pp; English.  
 CC  
 CC The invention relates to isolated nucleic acid molecules and their  
 CC encoded secreted proteins. The nucleic acids and proteins are used to  
 CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,  
 CC rabbits, goats, horses, cats, dogs, chickens or sheep. They  
 CC are also used in diagnosing a pathological condition or susceptibility  
 CC to a pathological condition. Antibodies to the proteins can also

CC be used in alleviating symptoms associated with the disorders and in  
 CC diagnostic immunoassays e.g. radioimmunoassays or enzyme linked  
 CC immunoassay assays (ELISA). Disorders which are diagnosed or treated  
 CC include autoimmune diseases e.g. rheumatoid arthritis,  
 CC hyperproliferative disorders e.g. neoplasms of the breast or liver,  
 CC cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders  
 CC e.g. cerebral ischemia, angiogenesis, nervous system disorders e.g.  
 CC Alzheimer's disease, infections caused by bacteria, viruses and fungi  
 CC and ocular disorders e.g. corneal infection. The polypeptides can also  
 CC be used to aid wound healing and epithelial cell proliferation, to  
 CC prevent skin aging due to sunburn, to maintain organs before  
 CC transplantation, for supporting cell culture of primary tissues, to  
 CC regenerate tissues and in chemotaxis. The polypeptides can also be used  
 CC as a food additive or preservative to increase or decrease storage  
 CC capabilities, fat content, lipid, protein, carbohydrate, vitamins,  
 CC minerals, cofactors and other nutritional components. The present

alignment\_scores:  
 Quality: 1720.00 Length: 317  
 Ratio: 5.426 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
 US-10-040-803-7 x AAS26871 ..  
 Align seg 1/1 to: AAS26871 from: 1 to: 1439

1 MetValValSerGlyAlaProProAlaLeuGlyGlyCysLeuGlyThr 17  
 ||||||||||||||||||||||||||||||||||||||||||||  
 104 ATGTGTGTTTCTGGAGCGCCCAACCCCTGGTGGGGCTCTCTCGGAC 153  
 17 PheThrSerLeuLeuLeuLeuAlaSerThrAlaIleLeuAlaIle 34  
 ||||||||||||||||||||||||||||||||||||||||||||  
 134 CTTCACCTCCCTGCTGCTGCTGGCGTCGACAGCATCTCAATCGGCCA 203  
 34 rGlieProValProProAlaCysGlyGlyProGlnGlnLeuAsnVal 50  
 ||||||||||||||||||||||||||||||||||||||||||||  
 204 GGATACCTGTTCCCACTGCTGGGAGCCCAAGCACTGAACCGGATT 253  
 51 ValGlyGlyGluAspSerThrAspSerGluTrpProTrpIleValSer 67  
 ||||||||||||||||||||||||||||||||||||||||||||  
 254 GTGGGGCGGCGAGACAGCAGCAGCAGCAGCGGCGCTGAGACAT 303  
 67 eGlnLysAsnGlyThrHisLysCysAlaGlySerLeuLeuThrSerArg 84  
 ||||||||||||||||||||||||||||||||||||||||||||  
 304 CCAGAAAGATGGAGCCACCACTGCGAGCTTCTGTCACCGCCGCT 353  
 84 rPValIleThrAlaAlaHisCysPheLysAspAsnLeuAsnLysProTyr 100  
 ||||||||||||||||||||||||||||||||||||||||||||  
 354 GGATGATCACTGCTGCCCACTGTTTCAAGGACAAACCTGAACAACTATC 403  
 101 LeuPheSerValLeuLeuGlyAlaTrpPheLeuGlnLysProGlySerAr 117  
 ||||||||||||||||||||||||||||||||||||||||||||  
 404 CTGTTCTCTGTGCTGCTGGGGCCCTGGCAGCTGGGGAACCTGGCTCTCG 453  
 117 gSerGlnLysValGlyAlaIleTrpValGluProHisProValTyrSerT 134  
 ||||||||||||||||||||||||||||||||||||||||||||  
 454 GTCCCAAGAGGTGGGTGTTGCTGGGTGAGCCCACTGTTATTCCT 503  
 134 rPlyGlnGlyAlaCysAlaAspIleAlaLeuValArgLeuGlnArgSer 150  
 ||||||||||||||||||||||||||||||||||||||||||||  
 504 GGAGGAAAGGTGCTGTCAGACATTTGCCCTGGTGCTGAGCGCTCC 553  
 151 IleGlnPheSerGluArgValLeuProIleCysLeuProAspAlaSerIle 167  
 ||||||||||||||||||||||||||||||||||||||||||||  
 554 ATACAGTTCAGAGCGGGCTCTGCCCATCTGCTACCTGATGCTCTCAT 603  
 167 eHisLeuProProAsnThrHisCysTrpPheSerGlyTrpGlySerIleG 184  
 ||||||||||||||||||||||||||||||||||||||||||||  
 604 CCACCTCCCTCAAAACCACTGCTGATCTCAGCTGGGGGAGCATTC 653  
 184 InAspGlyValProLeuProHisProGlnThrLeuGlnLysLeuVal 200

PR	07-JUL-2000;	2000US-0216680
PR	11-JUL-2000;	2000US-0217487
PR	11-JUL-2000;	2000US-0217496
PR	14-JUL-2000;	2000US-0218290
PR	14-JUL-2000;	2000US-0220963
PR	26-JUL-2000;	2000US-0220964
PR	26-JUL-2000;	2000US-0224518
PR	14-AUG-2000;	2000US-0224519
PR	14-AUG-2000;	2000US-0225213
PR	14-AUG-2000;	2000US-0225266
PR	14-AUG-2000;	2000US-0225267
PR	14-AUG-2000;	2000US-0225268
PR	14-AUG-2000;	2000US-0225270
PR	14-AUG-2000;	2000US-0225547
PR	14-AUG-2000;	2000US-0225575
PR	14-AUG-2000;	2000US-0225758
PR	14-AUG-2000;	2000US-0225759
PR	18-AUG-2000;	2000US-0226279
PR	18-AUG-2000;	2000US-0226661
PR	22-AUG-2000;	2000US-0226868
PR	22-AUG-2000;	2000US-0227182
PR	22-AUG-2000;	2000US-0227709
PR	30-AUG-2000;	2000US-0228924
PR	01-SEP-2000;	2000US-0229287
PR	01-SEP-2000;	2000US-0229343
PR	01-SEP-2000;	2000US-0229344
PR	01-SEP-2000;	2000US-0229345
PR	05-SEP-2000;	2000US-0229509
PR	05-SEP-2000;	2000US-0229513
PR	06-SEP-2000;	2000US-0230437
PR	06-SEP-2000;	2000US-0230438
PR	06-SEP-2000;	2000US-0231242
PR	08-SEP-2000;	2000US-0231243
PR	08-SEP-2000;	2000US-0231244
PR	08-SEP-2000;	2000US-0231413
PR	08-SEP-2000;	2000US-0231414
PR	08-SEP-2000;	2000US-0232080
PR	08-SEP-2000;	2000US-0233081
PR	12-SEP-2000;	2000US-0231968
PR	14-SEP-2000;	2000US-0233397
PR	14-SEP-2000;	2000US-0233398
PR	14-SEP-2000;	2000US-0233399
PR	14-SEP-2000;	2000US-0232400
PR	14-SEP-2000;	2000US-0234401
PR	14-SEP-2000;	2000US-0234402
PR	14-SEP-2000;	2000US-0233064
PR	14-SEP-2000;	2000US-0233065
PR	14-SEP-2000;	2000US-0234223
PR	21-SEP-2000;	2000US-0234274
PR	25-SEP-2000;	2000US-0234977
PR	25-SEP-2000;	2000US-0234998
PR	25-SEP-2000;	2000US-0235834
PR	26-SEP-2000;	2000US-0235834
PR	27-SEP-2000;	2000US-0235836
PR	27-SEP-2000;	2000US-0235837
PR	29-SEP-2000;	2000US-0236367
PR	29-SEP-2000;	2000US-0236368
PR	29-SEP-2000;	2000US-0236369
PR	29-SEP-2000;	2000US-0236370
PR	29-SEP-2000;	2000US-0236802
PR	02-OCT-2000;	2000US-0237037
PR	02-OCT-2000;	2000US-0237038
PR	02-OCT-2000;	2000US-0237039
PR	02-OCT-2000;	2000US-0237040
PR	13-OCT-2000;	2000US-0239353
PR	13-OCT-2000;	2000US-0239357
PR	20-OCT-2000;	2000US-0240960
PR	20-OCT-2000;	2000US-0241221
PR	20-OCT-2000;	2000US-0241785
PR	20-OCT-2000;	2000US-0241786
PR	20-OCT-2000;	2000US-0241808



201 ProtlelleAspsergluValCysSerHisLeuTyrTrpArgLysAla 217  
|||||  
617 CCTATCATGACTCGAAGTCTGACCATCTGTACTGGGGGAGACAG 666  
217 YGlnGlyProIleThrGluAspMetLeuCysAlaGlyTyrLeuGluGly 234  
|||||  
667 ACAGGACCATCATCTAGGACATGCTGTGGCGGCTACTTGAGAGGG 716  
234 LuArgAspAlaCysLeuGlyAspserGlyProLeuMetCysGluVal 250  
|||||  
717 AGCGGAGATCTGTCTGGCGACATCCGGGGCCCCCTCATGTGCGAGTG 766  
251 AspGlyAlaTrpLeuLeuAlaGlyIleIleSerTrpGlyGluGlyCysAl 267  
|||||  
767 GAGGGCGCTGCTGCTGCTGGCGATCATCGTGGGGGCGAGGGCTGTGC 816  
267 aGluArgAsnArgProGlyValTyrIleSerLeuSerAlaHisArgSer 284  
|||||  
817 CGAGGCAACAGCCCGGGGTCTACATCATGCTCTGCGCGACGCTCTCT 866  
284 rValGluIleTyrIleValGlnGlyValGlnLeuArgGlyArgAlaGlnGly 300  
|||||  
867 GGCTGAGAGATGCTGCAAGGGGTGCGAGCTCCGGGGCGGCTCAGAGG 916  
301 GLyGlyAlaLeuArgAlaProSerGlnGlySerGlyAlaAlaAlaArgse 317  
|||||  
917 GGTGGGGCCCTCAGGAGCCAGCCAGGCTCTGTGGGGCGGCGCGCTC 966  
317 r 317  
|  
967 C 967  
seq\_name: /SIDS1/gcdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAS26943  
seq\_documentation\_block:  
ID AAS26943 standard; CDNA: 1352 BP.  
XX  
AC AAS26943;  
XX  
DT 07-NOV-2001 (first entry)  
XX  
DE Human cDNA encoding a novel secreted protein, SEQ ID 135.  
XX  
XX Human; immunosuppressive; antiarthritic; ss; antirheumatic;  
cytostatic; cardiant; vasotropic; cerebroprotective; neurotropic;  
neuroprotective; antibacterial; virucide; fungicide; ophthalmological;  
vulnary; secreted protein; rheumatoid arthritis;  
hyperproliferative disorder; cardiovascular disorder; cardiac arrest;  
cerebrovascular disorder; cerebral ischemia; angiogenesis;  
nervous system disorder; Alzheimer's disease; infection; ocular disorder;  
corneal infection; wound healing; epithelial cell proliferation;  
skin aging; food additive; preservative; antiproliferative.  
XX  
OS Homo sapiens.  
XX  
PN WO200155441-A2.  
XX  
PD 02-AUG-2001.  
XX  
PF 17-JAN-2001; 2001WO-US01320.  
XX  
PR 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.

PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 14-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226686.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235835.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239335.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.

be used in alleviating symptoms associated with the disorders and in diagnostic immunoassays e.g. radioimmunoassays or enzyme linked immunosorbent assays (ELISA). Disorders which are diagnosed or treated include autoimmune diseases e.g. Rheumatoid arthritis, hyperproliferative disorders e.g. neoplasms of the breast or liver, cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorder e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g. Alzheimer's disease, infections caused by bacteria, viruses and fungi and ocular disorders e.g. corneal infection, and many other disorders listed in the specification. The polypeptides can also be used to aid wound healing and epithelial cell proliferation, to prevent skin aging due to sunburn, to maintain organs before transplantation, for supporting cell culture of primary tissues, to regenerate tissues and in chemotaxis. The polypeptides can also be used as a food additive or preservative to increase or decrease storage capabilities, fat content, lipid, protein, carbohydrate, vitamins, minerals, cofactors and other nutritional components. The present

```

alignment_scores:
    Quality: 1714.00      Length: 317
    Ratio: 5.424          Caps: 0
Percent Similarity: 99.685   Percent Identity: 99.685

Alignment_block:
US-10-040-803-7 x AAS26943 ..

Align seg 1/1 to: AAS26943 from: 1 to: 1352

1 MetValValSerGlyAlaProProAlaLeuGlyGlyCysLeuGlyIth 17
|||||
17 ATGGTGGTTTCGTGAGCGCCCGACGCCCTGGGGGGGCTGTCTGCAC 66

17 rPhetThSerLeuLeuLeuLeuAlaSerThrAlaIleLeuSnpAlaIaa 34
|||||
67 CTTCACCTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCA 116

34 rgllrProValProProAlaCysGlyLysProGlnGlnLeuAsnArgVal 50
|||||
117 GGATACCTGTTCCCCCAGCGCTGGGAAAGCCCGAACCTGAACGGGGTT 166

51 ValGlyGlyLysAspSerThrAspSerGluTrpProTrpIleValSerI 67
|||||
167 GTGGCGGCGAGGACAGCACACTACAGCCAAGTGCCCTGGACCTGACCAT 216

67 egInLyAsnGlyThrHisHisCysAlaGlySerLeuLeuThrSerArg 84
|||||
217 CKAAGAATAAGGAGCCACCACACTGGGAGAGTTCTCTGTACCAAGCCGT 266

84 rPvalIleThrAlaAlaHisCysPheLysAspAsnLeuAsnLysProTry 100
|||||
267 GGGGTGATCACTGCTGCCCACTGTTCAGGACAACCTGAACAAACAATAC 316

101 LeuPheSerValLeuLeuGlyAlaAlaTrpGlnLeuLysAsnProGlySer 117
|||||
317 CTGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 366

117 gSerGlnLysValGlyAlaAlaTrpValGluProHisProValTrySerT 134
|||||
367 GTCCCAAGAGTGGGGTGGCTGGGTGGAGGCCCAACCTGTATTCTCT 416

134 rPyLysGlnGlyAlaCysAlaAspIleAlaLeuValArgLeuGlnArgSer 150
|||||
417 GGAAGAAAGTGCCTGTGCAGACATTCCTCCCTGGCTGCTGCCAGCCCTCC 466

151 IleGlnPheSerGluArgValLeuProIleCysLeuProAspAlaSerI 167
|||||
467 ATACAGATTCTCAGAGCGGGCTCTCCCATCTGGCTTAACCTGATCCCTAT 516

167 ehIstLeuProAsnThrHisCysTrpIleSerGlyTrpGlySerIleG 184
|||||
517 CCACCTCCCTCCAAAACCCCACTGCTGATCTCAGGCTGGGGAGAGATCC 566

184 InAsnGlyValProLeuProHisProGlyTrpThrGlnGlyValValVal

```

```
|||||
567 AAGATGAGTTCCTTGCCCAACCCAGACCCCTGAGAGCTGAAGTT 616
|||||
201 ProIleIleAspSerGluValCysSerHisLeuYrTTPArgIylAgl 217
|||||
617 CCTATCATGACTCGGAGCTGCAAGCCATCTGTACTGCGGGGAGCAGG 666
|||||
217 yGInGIyProIleThrgLuAspMetLeuCysAglYrTLeuGIuGIyG 234
|||||
667 ACAGGAGCCCATCCTACTGAGGACATGCTGTGTCCGCTACTTGGAGGGGG 716
|||||
234 luArgAspAlaCysLeuGIyAspSerGIyGIyProLeuMetCysGlnVal 250
|||||
717 AGCGGATGCTGTGCTGGGCGACTCCGGGGGCCCTCATGTGCCAGGTG 766
|||||
251 AspGIyAlaTrpLeuLeuAglYrIleSerTPGIyGIuGIyCysAl 267
|||||
767 GACGGGCGCTGGCTGCTGGCCGCATCATCAGCTGGGGCGAGGCTGTGC 816
|||||
267 agluArgAsnArgProGIyValYrIleSerLeuSerAlaHisArgSerT 284
|||||
817 CGAGGCGCAACAGGCCCGGGGTCTACATCAGCCTCTGTGGCACCCTCTCT 866
|||||
284 rPValGIuLysIleValGIuGIyValGlnLeuArgGIyArgAlaGIuGIy 300
|||||
867 GGTGGAGAGATCGTGCAGGGGTGCAGCTCCGGGGCGGCGCTCAGGGG 916
|||||
301 GIyGIyAlaLeuArgAlaProSerGIuGIySerGIyAlaAlaAlaArgSe 317
|||||
917 GGTGGGGCCCTCAGGGCAACCGAGCGAGGCTCTGGGGCGCGCGCGCTC 966
|||||
317 r 317
|
967 C 967
```





OM of: US-10-040-803-7 to: EST:\* out\_format: pfs  
Date: Aug 13, 2002 1:04 PM

About: Results were produced by the Gencore software, version 4.5.  
Copyright (c) 1993-2000 Compugen Ltd.

## Command line parameters:

-MODEL=frame+ p2n.model -DEV=x11  
-O=/cpn2.1/US10040803/runat.13082002.083351.7327/app-query.fasta.1.378  
-DB=EST -QPM=fastap -SUFFIX=p2n.rst -GAPOP=12.000 -GAPEXT=4.500  
-MIMATCH=0.100 -LOOPEXT=0.000 -LOOPEXT=0.000 -GAPOP=4.500  
-OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000  
-FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DEL0P=6.000  
-DEL0P=7.000 -START=1 -MATRIX=blomsum62 -TRANS=human40.cdi  
-LIST=45 -DOCLIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0  
-ALIGN=10 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -HEAPSIZE=500  
-MULEN=0 -MAXLEN=2000000000 -USFR=US10040803.cgnl.1.2673  
-D=6 -ICPU=3 -LONGLOG -DEV\_TIMEOUT=120 -MAIN\_TIMEOUT=30  
-ALPXY -WALT -THREADS=1

## Query information block:

Query: US-10-040-803-7  
Query length: 317  
Database: EST\*  
Database sequences: 13736207  
Database length: -184157050  
Search time (sec): 1684.510000

## Score\_list:

Sequence	Strd	Orig	ZScore	Escore	Len	Documentation
gb_hic:AK014645	+	1271.00	1292.87	9.3e-102	1323	AK014645 Mus musculus 10 days
gb_est2:BE907706	+	812.00	1277.97	6.2e-62	470	BE907706 60197734F1 NIH_MGC_70
gb_est2:BG824600	+	804.00	1260.82	5.6e-61	712	BG824600 602728529F1 NIH_MGC_15
gb_est2:BG824236	+	732.00	1151.35	7.0e-55	473	AM462236 BP230008B20C12 Soares
gb_est2:BF180373	+	710.50	1109.97	1.4e-51	935	BF180373 601804557F1 NCI_CGAP_M
gb_est1:AM463601	+	692.00	1088.62	2.2e-51	450	AM463601 BP230013A20B9 Soares
gb_est1:AM463315	+	657.00	1031.96	3.1e-46	507	AM463315 PM1-HT0345-161199-002
gb_est1:AM463315	+	600.00	946.44	1.8e-43	703	AM463315 IL2-CT0031-290799-001
gb_est2:BM389391	+	586.00	916.11	8.8e-42	703	BM389391 UT-R-CN1-G1-1-03-0-01
gb_est2:BG480197	+	580.50	903.58	4.4e-41	1001	BG480197 602530380F1 NIH_MGC_2
gb_est2:BG679626	+	569.00	894.79	1.4e-40	420	AM462598 BP230009B20E11 Soares
gb_est2:BG679626	+	565.00	879.66	9.4e-40	947	BG679626 602627678F1 NCI_CGAP_5
gb_est1:AM51470	+	564.50	878.29	1.1e-39	999	AL551470 AL551470 LFI_NFL006.PI
gb_est1:AM51470	+	561.50	883.32	5.9e-40	405	AM51470 LFI_NFL006.PI
gb_est1:AM51470	+	542.50	838.20	1.9e-37	1629	AK010640 Mus musculus ES cells
gb_est1:AM51470	+	540.00	827.52	7.6e-37	3030	AK004939 Mus musculus adult ma
gb_est1:AM51470	+	532.00	826.34	8.8e-37	1050	AK006271 Mus musculus adult ma
gb_est1:AM51470	+	521.50	811.69	7.0e-36	876	AL55870 AL55870 LFI_NFL006.PI
gb_est1:AM51470	+	521.00	810.18	7.0e-36	936	AL55870 AL55870 LFI_NFL006.PI
gb_est1:AM51470	+	517.50	806.32	1.1e-35	802	BG962187 60287694F1 NCI_CGAP_C
gb_est1:AM51470	+	517.00	813.76	4.4e-36	375	AM383297 PM1-HT0345-121199-001
gb_est1:AM51470	+	508.00	788.99	1.1e-34	1000	BI13250 60287694F1 NCI_CGAP_C
gb_est1:AM51470	+	507.00	788.59	1.1e-34	889	BI13250 602937954F1 NCI_CGAP_C
gb_est1:AM51470	+	500.00	777.87	4.4e-34	861	BI13250 602937954F1 NCI_CGAP_C
gb_est1:AM51470	+	496.00	769.86	1.2e-33	1006	BI13250 602937954F1 NCI_CGAP_C
gb_est1:AM51470	+	490.50	767.63	1.6e-33	553	BM389391 PM1-HT0345-121199-001
gb_est1:AM51470	+	480.50	764.16	2.4e-33	728	BM389391 PM1-HT0345-121199-001
gb_est1:AM51470	+	480.50	755.13	2.4e-33	1052	BI13250 60287694F1 NCI_CGAP_C
gb_est1:AM51470	+	480.50	753.33	1.0e-32	745	BI13250 60287694F1 NCI_CGAP_C
gb_est1:AM51470	+	480.00	743.63	3.5e-32	1089	AK016509 Mus musculus adult ma
gb_est1:AM51470	+	478.50	747.93	2.1e-32	591	BE907706 60197734F1 NIH_MGC_70
gb_est1:AM51470	+	473.00	736.27	9.2e-32	777	BI259237 602970095F1 NIH_MGC_12
gb_est1:AM51470	+	473.00	724.78	4.0e-31	2246	AK005546 Mus musculus adult fe
gb_est1:AM51470	+	469.50	719.17	7.6e-31	426	BM389391 PM1-HT0345-121199-001
gb_est1:AM51470	+	459.50	711.15	2.3e-30	821	BM389391 PM1-HT0345-121199-001
gb_est1:AM51470	+	457.50	710.92	2.5e-30	862	BI1344893 60299736F1 NCI_CGAP_C
gb_est1:AM51470	+	454.00	704.97	5.1e-30	871	BI1344893 60299736F1 NCI_CGAP_C
gb_est1:AM51470	+	453.00	705.16	5.0e-30	780	AM202972 f1240404.y1 zebrafish a
gb_est1:AM51470	+	451.50	705.27	4.9e-30	588	BI074421 BI074421 NIBB Mochi f
gb_est1:AM51470	+	451.50	700.68	8.8e-30	899	BM389391 PM1-HT0345-121199-001

seq\_name: gb\_hic:AK014645  
seq\_documentation block:  
LOCUS AK014645 1323 bp mRNA linear HTC 19-JAN-2002  
DEFINITION Mus musculus 10 days neonate skin cDNA, RIKEN full-length enriched library, clone:47334401N09:BRAIN SPECIFIC SERINE PROTEINASE-4 TYPE1, full insert sequence.  
ACCESSION AK014645  
VERSION AK014645.1 GI:12852632  
KEYWORDS HTC; CAP trapper.  
SOURCE Mus musculus (strain:C57BL/6J) 10 days neonate skin cDNA to mRNA, clone:47334401N09.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

REFERENCE 1 (sites)  
Carninci, P. and Hayashizaki, Y.  
High-efficiency full-length cDNA cloning  
Meth. Enzymol. 303, 19-44 (1999)  
JOURNAL 99279253  
PUBMED 10349636  
2 (sites)  
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes  
Genome Res. 10 (10), 1617-1630 (2000)  
JOURNAL 20499374  
PUBMED 11042159  
3 (sites)  
Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Kono, H., Akiyama, J., Nishi, K., Kitsuana, T., Tashiro, H., Itoh, M., Suni, N., Ishii, Y., Nakamura, S., Hazama, M., Nishibe, T., Harada, A., Yamamoto, R., Matsunoto, H., Sakauchi, S., Ikegami, T., Kashiwagi, K., Fuyukawa, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watabiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A., and Hayashizaki, Y.  
RIKEN integrated sequence analysis (RISA) system -384-format sequencing pipeline with 384 multicapillary sequencer  
Genome Res. 10 (11), 1757-1771 (2000)  
JOURNAL 20530913  
PUBMED 11076861  
4 (sites)  
FANTOM Consortium.  
Functional annotation of a full-length mouse cDNA collection  
Nature 409, 685-690 (2001)  
JOURNAL 5 (bases 1 to 1323)

ADACHI, J., Aizawa, K., Akahira, S., Akimura, T., Aono, H., Arai, A., Arakawa, T., Balderelli, R., Bono, H., Brownstein, M., Bull, C., Carinini, P., Fukuda, S., Fukunishi, Y., Furuno, M., Hanaagaki, T., Hara, A., Hayatsu, N., Hill, D., Hiramoto, K., Hiraoka, T., Hori, F., Hume, D., Imotani, K., Ishii, Y., Itoh, M., Izawa, M., Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Kono, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Okazaki, Y., Okido, T., Owa, C., Quackenbush, J., Saito, H., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D., Shimizu, L., Shibata, K., Shibata, Y., Shingawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Tanaka, T., Tejima, Y., Toyota, T., Yamamura, T., Yamakawa, I., Yasunishi, A., Yoshida, K., Yoshino, M., Muramatsu, M., and Hayashizaki, Y.  
Direct Submission  
Submitted (10-JUN-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC),



CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNL at:  
<http://image.llnl.gov>  
 Plate: LLM9697 row: d column: 09  
 High quality sequence stop: 468.  
 Location/Qualifiers

## FEATURES

source

1..470  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:3899480"  
 /clone\_lib="NIH\_MGC\_70"  
 /tissue\_type="epithelioid carcinoma"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Organ: pancreas; Vector: pCMV-Sport6; Site\_1: NotI;  
 Site\_2: SalI; Cloned unidirectionally. Primer: Oligo dT.  
 Average insert size 1.1 kb. Library constructed by Life  
 Technologies"

COUNT 77 a 155 c 145 g 93 t

## alignment\_scores:

Quality: 812.00 Length: 151  
 Ratio: 5.377 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

## alignment block:

US-10-040-803-7 x BE907706

Align seg 1/1 to: BE907706 from: 1 to: 470

```

1 MetValValSerGlyAlaProProAlaLeuGlyGlyCysLeuGlyTh 17
|||||
18 ATGGTGGTTTCTGGAGCGCCCGACCGCTGGGTGGGGCTGTCGGCAC 67
|||||
17 rphenHsrLeuLeuLeuAlaSerThrAlaIleLeuSnaAlaAla 34
|||||
68 CTTCACCTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 117
|||||
34 rGleProValProProAlaCysGlyLysProGlnGlnLeuSnaArgVal 50
|||||
118 GGATACCTTTCCCGACGCTGTGGAGCCCGACACCTGTAACCGGGTT 167
|||||
51 ValGlyGlyLysAspSerThrAspSerGluTrpProTrpIleValSer 67
|||||
68 GTGGGGCGGAGAGACAGCACTGACAGCGAGTGGCCCTGATCGTAGCAT 217
|||||
67 eGlnLysAsnGlyThrHisHisCysAlaGlySerLeuLeuThrSerArg 84
|||||
218 CCAGAAAGATGGGACCCACCACTGCGCAGTCTCTGCTCACACCGCCCT 267
|||||
84 rValIleThrAlaAlaHisCysPheLysAspAsnLeuSnaLysProTrp 100
|||||
268 GGGTGAFTCACTGCTGCCACTGTTCAAGGACAACCTGAAACCAATAC 317
|||||
101 LeuPheSerValLeuLeuGlyAlaTrpGlnLeuGlyAsnProGlySerAr 117
|||||
318 CTGTCTCTGTGCTGCTGGGGGCTGCGACGCTGGGGAAACCTGGCTCTCG 367
|||||
117 gSerGlnLysValGlyValAlaTrpValGluProHisProValIlySer 134
|||||
368 GTCCCAAGAGGTGGGTGGCTGGGGTGAAGCCCAACCTGTGATTCCT 417
|||||
134 rPlySerGluGlyAlaCysAlaAspIleAlaLeuValArgLeuGlnArgSer 150
|||||
418 GGAAGAAAGTGGCTGTGACAGATTGCCCTGCTGCTGCTGAGGCTCC 467
|||||
151 Ile 151
|||||
468 ATA 470

```

seq\_name: gb\_est2:BG824600

seq\_documentation\_block:

LOCUS

BG824600

712 bp

mRNA

linear EST 22-MAY-2001

DEFINITION

602728529F1 NIH\_MGC\_15 Homo sapiens cDNA clone IMAGE:4868047 5',

ACCESSION

BG824600

mRNA sequence.

VERSION

BG824600.1

GI:14172187

KEYWORDS

EST.

SOURCE

human.

ORGANISM

Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE

1 (bases 1 to 712)

AUTHORS

NIH-MGC <http://imgc.nci.nih.gov/>.

TITLE

National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL

Unpublished (1999)

COMMENT

Contact: Robert Strausberg, Ph.D.  
 Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)

Tissue Procurement: ATCC

CDNA Library Preparation: Ling Hong/Rubin Laboratory

DNA Sequencing by: NIH Intramural Sequencing Center

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LNL at:

<http://image.llnl.gov>

Plate: LLM1737 row: 1 column: 08

High quality sequence stop: 695.

Location/Qualifiers

1..712

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:4868047"

/clone\_lib="NIH\_MGC\_15"

/tissue\_type="adenocarcinoma cell line"

/lab\_host="DH10B (phage-resistant)"

/note="Organ: colon; Vector: pOB7; Site\_1: XhoI; Site\_2:

EcoRI; cDNA made by oligo-dT priming. Directionally

cloned into EcoRI/XhoI sites using the following 5'

adaptor: GGCGAGGAG(c). Size-selected &gt;500bp for average

insert size 1.8kb. Library constructed by Ling Hong in

the laboratory of Gerald M. Rubin (University of

California, Berkeley) using ZAP-cDNA synthesis kit

(Stratagene) and Superscript II RT (Life Technologies)"

BASE COUNT

109 a

234 c

240 g

129 t

ORIGIN

alignment\_scores:

Quality: 804.00 Length: 317  
 Ratio: 4.814 Gaps: 1  
 Percent Similarity: 52.681 Percent Identity: 52.681

## alignment block:

US-10-040-803-7 x BG824600

Align seg 1/1 to: BG824600 from: 1 to: 712

```

1 MetValValSerGlyAlaProProAlaLeuGlyGlyCysLeuGlyTh 17
|||||
55 ATGGTGGTTTCTGGAGCGCCCGACCGCTGGGTGGGGCTGTCGGCAC 104
|||||
17 rphenHsrLeuLeuLeuAlaSerThrAlaIleLeuSnaAlaAla 34
|||||
105 CTTCACCTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 154
|||||
34 rGleProValProProAlaCysGlyLysProGlnGlnLeuSnaArgVal 50
|||||
155 GGATACCTTTCCCGACGCTGTGGAGATCCCGACAGTGAACCGGGTT 204
|||||
51 ValGlyGlyLysAspSerThrAspSerGluTrpProTrpIleValSer 67
|||||
205 GTGGGGCGGAGAGACAGCACTGACAGCGAGTGGCCCTGATGTGATCAT 254

```

67 eGlnIlysaSngIyThrhiNhiScysAlaGlySerLeuLeuThiSerArgT 84  
 |||||  
 255 CCAAGAAATGGGACCCACACAGTGGAGTTCTGTGTCACCAAGCCGCT 304  
 84 rPvAlIleThrAlaIAlaIhiScysPheIysAspAsnLeuAsnIlyProTyr 100  
 |||||  
 305 GGGGATCATCTAGTGGCCACATGTTTCAAGGACCAACCTGAAACCAACATAC 354  
 101 LeuPheSerValLeuLeuGlyAlaTrpGlnIleuGlyAsnProGlySerArg 117  
 |||||  
 355 CTGTTCCTCTGTGCTGCTGGGGCCCTGGCAGCTGGGAAACCTGGCTCTCG 404  
 117 gSerGlnIlyValGlyAlaIAlaTrpValGlnProhiSerProValTyrSerT 134  
 |||||  
 405 GTCCCAAAAGGTGGGTGTTCCTGGGTGAGGCCACCCCTGTGTAT ... 450  
 134 rPlySGlnGlyAlaCysAlaAspIleAlaIleuValArgLeuGlnIlySer 150  
 450 ..... 450  
 51 IleGlnPheSerGluArgValLeuProIleCysLeuProAspAlaSerI 167  
 450 ..... 450  
 167 enIleuProProAsnThiNhiScysTrpIleSerGlyTrpGlySerIleG 184  
 450 ..... 450  
 184 IAspArgIValProLeuProhiSerProGlnThiLeuGlnIlyLeuIlyVal 200  
 450 ..... 450  
 201 ProIleIleAspSerGluValCysSerhiNhiLeuTyrTrpArgGlyAlaG 217  
 450 ..... 450  
 217 yGlnGlyProIleThiGlnAspMetLeuCysAlaGlyTyrLeuGlnIlyG 234  
 450 ..... 450  
 234 IuArgAspAlaCysLeuGlyAspSerGlyIlyProLeuMetCysGlnVal 250  
 450 ..... 450  
 251 AspGlyAlaTrpLeuLeuAlaGlyIleIleSerTrpGlyGlnIlyCysAl 267  
 450 ..... 450  
 267 agIuArgAsnArgProGlyValTyrIleSerLeuSerAlaIhiArgSerT 284  
 451 ..... TCTT 454  
 284 rPvAlGlnIlyIleValGlnIlyValGlnIleuArgGlyArgAlaGlnIly 300  
 |||||  
 455 GGGTGAAGAAGATCGTCAAGGGGTGACGCTCCGGGGCCGCTCAAGGG 504  
 301 GlnGlyAlaLeuArgAlaProSerGlnIlySerGlyAlaIAlaIAlaArgse 317  
 |||||  
 505 GGTGGGGCCCTCAGGGACCGCAAGGGCTGTGGGGCCGCGCGGCTC 554  
 317 r 317  
 555 C 555  
 seq\_name: gb\_est1:AW462236  
 seq\_documentation\_block:  
 LOCUS AW462236 473 bp mRNA linear EST 24-FEB-2000  
 DEFINITION BP230008B20C12 Soares normalized bovine placenta Bos taurus cDNA  
 clone BP230008B20C12 5', mRNA sequence.  
 ACCESSION AW462236  
 VERSION AW462236.1 GI:7032404

KEYWORDS EST.  
SOURCE cow.  
ORGANISM Bos taurus  
Euteleostomi;  
Mammalia; Euteria; Chordata; Craniata; Vertebrata; Euteleostomi;  
Bovidae; Bovinae; Bos.  
REFERENCE 1 (bases 1 to 473)  
AUTHORS Lewin, H.A., Soares, M.B., Rebelz, M., Pardinas, J., Liu, L. and Larson, J. H.

TITLE Bovine ESTs  
JOURNAL Unpublished (2000)  
COMMENT Contact: Lewin, H. A.  
W. M. Keck Center for Comparative and Functional Genomics  
University of Illinois at Urbana-Champaign  
340 Edward R. Madigan Laboratory, 1201 W. Gregory Dr., Urbana, IL  
61801, USA  
Tel: 217 244 5998  
Fax: 217 244 5617  
Email: h-lewin@uiuc.edu

Funding for cattle EST sequencing was provided by the USDA National Research Initiative, Animal Genome Resource Grant AG 99-3205-8534 to H. A. Lewin and J. E. Womack. Base Calling/Quality Scores: PHRED from Washington University Genome Center. Vector Trimml g:  
Cross-match from Washington University Genome Center PHRAP suite.  
Sequences submitted are vector free and at least 200 bp in length.  
PCR Primers  
FORWARD: TAATACGACTCATATAGGG  
BACKWARD: ATATACCGTCACCTAAG  
Insert Length: 473 Std Error: 0.00  
Plate: BP230008B20 row: C column: 12  
Seq primer: AGCGGATACCAATTTCACACGGA  
High quality sequence stop: 473.  
Location/Qualifiers  
1. .473  
/organism="Bos taurus"  
/db\_xref="taxon:9913"  
/clone="BP230008B20C12"  
/clone\_lib="Soares normalized bovine placenta"  
/sex="Female"  
/lab\_host="DH10B"  
/note="Organ: placenta; Vector: pT73pac; Site: 1: EcoRI;  
Site: 2: NotI; The cDNA library was contributed by the  
Soares laboratory and it was constructed and normalized  
as described by Bonaldo, M.F., Lennon, G. and Soares,  
M.B. (1996), Genome Research 6(9): 791-806."

BASE COUNT 79 a 178 c 138 g 78 t  
ORIGIN

alignment\_scores:  
Quality: 732.00 Length: 157  
Ratio: 4.913 Gaps: 0  
Percent Similarity: 94.904 Percent Identity: 82.166

alignment\_block:  
US-10-040-803-7 x AW462236 ..

Align seg 1/1 to: AW462236 from: 1 to: 473

66 Ser11IeglnlyASnglyThrHisHiscysAlaGlySerLeuIeuThrSe 82  
|||||  
3 AGCATCCGGAGAAACGGACACCCACACCGCGCGCTCCGTCACACAG 52  
82 rArgTrpAlIeThrAlaAlaHisCysPheIlyAspAsnuIeuAsnlysp 99  
|||||  
53 CCGCTGGGTGCTCACGGCCGCCACTGCTTAAAGATATATGTGACAAAC 102  
99 rGlyLeuPheSerValLeuIeuGlyAlaTrpGlnLeuGlyAsnPrGly 115  
|||||  
103 CAACCACTTCCTGTGCTGCTGGAGAGCTGGACGCTGGGAAACCGTGC 152  
116 SerArgSerGlnlyValGlyAlaAlaTrpValGluProHisProValTy 132  
|||||

```

153 CCAAGGTCCAGAGGTGGATATGCTGGGACAGCCCACTGTGTA 202
132 rsertrpysgluylalacysalaasplialeuvalargleu149
|||||
203 CTCCTGGAGAGGGCTCCCGGACATCGCTGGCTGGCTGGAGC 252
149 rsertrleu1pserglu1arqvalleuprolleu1cysleu1paspala 165
|||||
253 GCGCATTCAGTCTGTAGAGCGCTGCTGCGCATCTGCTGCGGACCTCC 302
166 sertrleu1pserglu1arqvalleuprolleu1cysleu1paspala 182
|||||
303 ACCGTCCAGCTCTCTCCGACACCAACTGATGCGGCTGGGAG 352
182 rleu1pserglu1arqvalleuprolleu1cysleu1paspala 199
|||||
353 CGTCCAGATGGAGTGCCTCCGACCTGACCTCCAGAGAGCTGA 402
199 ysvalprolleu1pserglu1arqvalleuprolleu1cysleu1paspala 215
|||||
403 AGGTCCCATCATGACACTGACGACCTGAGCGGCTGTACTGCGGGGA 452
216 AlaGlyGlyProLeuThr 222
|||||
453 GCCGGCAGGGCGCATACAC 473

```

seq\_name: gb\_est2:BF180373

seq\_documentation\_block:

LOCUS BF180373 935 bp mRNA EST 31-OCT-2000  
DEFINITION 601804557P1 NCI\_CGAP\_Mam5 Mus musculus cDNA clone IMAGE:4035351 5',  
mRNA sequence.  
ACCESSION BF180373  
VERSION BF180373.1 GI:11058515  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 935)  
AUTHORS NIH-MGC http://mgc.nci.nih.gov/  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: c9qpbs-r@mail.nih.gov  
Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys  
cDNA Library Preparation: Life Technologies, Inc.  
DNA Sequencing by: Inocyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNML at:  
http://image.llnl.gov  
Plate: LLM9309 row: a column: 16  
High quality sequence stop: 732.  
Location/Qualifiers  
1. 935

# FEATURES

source

/organism="Mus musculus"  
/strain="C57/B6"  
/db\_xref="taxon:10090"  
/clone="IMAGE:4035351"  
/clone\_lib="NCI\_CGAP\_Mam5"  
/tissue\_type="tumor, gross tissue"  
/dev\_stage="7 months"  
/lab\_host="DH10B"  
/notes="Organ: mammary; Vector: PCMV-SPORT6; Site:1: sal1;  
Site:2: NotI; Cloned unidirectionally. Primer: oligo dT.  
Library constructed by Life Technologies. Investigators  
providing samples: Lothar Hennighausen/Robin Humphreys,  
NIH"

BASE COUNT 197 a 279 c 287 g 172 t  
ORIGIN

alignment\_scores:  
Quality: 710.50 Length: 310  
Ratio: 3.230 Gaps: 9  
Percent Similarity: 70.968 Percent Identity: 55.806

alignment\_block:  
US-10-040-803-7 x BF180373

Align seg 1/1 to: BF180373 from: 1 to: 935

```

1 MetValIserGlyAlaProProAlaLeuGlyGlyCysLeuGlyTh 17
|||||
81 ATGATGATTCAGACACTCCCGACACTGGTGGGACGAGTTACAGAT 130
|||||
17 rphertrserleu1pserglu1arqvalleuprolleu1cysleu1paspala 34
|||||
131 CTAAATCCTTCTGTGCTGCTGATTCACAGACTCCATGAGTGTGCA 180
|||||
34 rgl1pserglu1arqvalleuprolleu1cysleu1paspala 50
|||||
181 CCATCCAGATGTCCTCCAGACTGTGGAGAGCTCCAGAGCTGACCGATT 230
|||||
51 ValGlyGlyGlyAlaPserThrAspSerGluTrpProTPIleValSer1 67
|||||
231 GTGGAGGTGAGACAGCATGATGCCAGTGGCTGGATGTTAGCAT 280
|||||
67 egluysasnglythrHisCysAlaGlySerLeuLeuThrSerArgT 84
|||||
281 CCTCAGAAATGGCTCCACACTGTGACAGCTCCCTGTCACCAACCTCT 330
|||||
84 rValIleThrAlaAlaHisCysPheIysAspAsnLeuAsnIysProTyr 100
|||||
331 GGTGTGTACAGCGCGCACTGCTTAAGACATATGACAAACCATCT 380
|||||
101 LeuPheSerValLeuLeuGlyAlaTrpGlnLeuGlyAsnProGlySerAr 117
|||||
381 CTGTTCACATATTGTTGGGGCTGGAGCTGGGAGCCAGCCCAAG 430
|||||
117 gSerGlnIysValGlyAlaAlaTrpValGluProHisProValTyrSerT 134
|||||
431 GTCCCAAAAGTAGGCAATGCTGGGTGCT .GCTCACCCAGCTATCTT 479
|||||
134 rPluGlyGlyAlaCysAlaAspIleAlaLeuValArgLeuGlnArgSer 150
|||||
480 GGAAGGAGGAGACCATGACAGATTCCTGCTGGCTGGAGACACCTCC 529
|||||
151 IleGlnPheSerGluArqValleuprolleu1cysleu1paspalaSer1 167
|||||
530 ATCCAGTCTGTAGCGGATCTGCCATCTGCTA .CTGACTCTCTCTT 578
|||||
167 eHisLeuProProAsnThrHisCysTrpIleSerGlyTTPGlySerIleG 184
|||||
579 CCGTCTCCCTCCCAAGACGAC .TGCTGGATTGGCG .TGGGGAAGCATCC 626
|||||
184 IuAspGlyValProLeuProHisProGlnThrLeuGlnIysLeuIysVal 200
|||||
627 AGGATGGGCTGCCCGGCC...ACCTCAGACTTCAGAAAGCGAGTGG 673
|||||
201 ProIleIleAspSerGluArqValleuprolleu1cysleu1paspalaSer1 217
|||||
674 CCA...TCCTGACTCGAAGCTGGAAGAGCTGTAC...CGGGGAGAGCG 717
|||||
217 yGlnGlyProIleThrGlnAspMetLeuCysAlaGlyTyrLeuGlnGlyG 234
|||||
718 GTCCAGAGCCATCACGGGCT...GCTGGTCTGCTGACTGGAG 758
|||||
224 IuArgAspAlaCysLeuGlyIAspSerGlyIysProLeuMetCysGlnVal 250
|||||
759 GAACGGAAGCGCTGTGGAGCT...GGGTCCTGAGTGTCCGGGTG 802
|||||
251 AspGlyAlaTrpLeuLeuAlaGlyIleIleSerTrpGlyIuGly..... 265
|||||
803 .....GCCGGCTGCGGAGACCGAGGAGGGGCGGAG 834

```

```

266 ... Cys1aagiuaargasnarpgrogiyValTyrilleseuleraleah 281
      ::::::::::::::::::::|
835 CCAGCCCGGGGACACCTAACCCGGGGAA..... 865
281 |aBgeSerrTpvaIgIuLysileValngiValgInLeuAargGlyAarg 297
      |||||
866 ..... GAGCAGGGG..... 874
298 AlaIngIyGlyGlyAlaLeuAargAlaPro 307
      ||::::|
875 GCCCGTGGGGGGGAAACATTAAGAACCCG 904

seq_name: gb_est1:AW463601

seq_documentation_block:
LOCUS      AW463601              450 bp      mRNA      linear      EST 24-FeB-2000
DEFINITION BP230013A20B9 Soares normalized bovine placenta Bos taurus cDNA
           clone BP230013A20B9 5', mRNA sequence.
VERSION    AW463601
XREFS      AW463601.1 GI:7033769
SOURCE     EST.
           cow.
ORGANISM   Bos taurus
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Cetartiodactyla; Kumlantia; Pecora; Bovidae;
           Bovidae; Bovinae; Bos.
REFERENCE   1 (bases 1 to 450)
           Lewin,H.A., Soares,M.B., Rebeiz,M., Pardinas,J., Liu,L. and Larson
           J.,H.
TITLE      Bovine ESTs
JOURNAL    Unpublished (2000)
COMMENT    Contract: Lewin, H. A.
           W. M. Keck Center for Comparative and Functional Genomics
           University of Illinois at Urbana-Champaign
           340 Edward R. Madigan Laboratory, 1201 W. Gregory Dr., Urbana, IL
           61801, USA
           Tel: 217 333 5998
           Fax: 217 244 5617
           Email: h-lewin@uiuc.edu
           Funding for cattle EST sequencing was provided by the USDA National
           Research Initiative, Animal Genome Resource Grant AG 99-3205-8534
           to H. A. Lewin and J. E. Womack. Base Calling/Quality Scores: PHRED
           from Washington University Genome Center. Vector Trimmi g:
           Cross_match from Washington University Genome Center PHRAP suite.
           Sequences submitted are vector free and at least 200 bp in length.
           PCR primers
           FORWARD: TAATACGACGACTATGACG
           BACKWARD: ATTAACGCTCAGTAAG
           Insert Length: 450 Std Error: 0.00
           Plate: BP230013A20 row: B column: 9
           Seq primer: AGCGATACCAATTCACACAGCA
           High quality sequence stop: 450.
           Location/Qualifiers
               1..450
               /organism="Bos taurus"
               /db_xref="taxon:9913"
               /clone="BP230013A20B9"
               /clone_id="Soares normalized bovine placenta"
               /sex="female"
               /lab_host="DH10B"
               /note="Organ: Placenta; Vector: pT733pac; Site:1: EcoRI;
               Site:2: NotI; The cDNA library was contributed by the
               Soares laboratory and it was constructed and normalized
               as described by Bonaldo, M.F., Lennon, G. and Soares,
               M.B. (1996), Genome Research 6(9): 791-806. "
BASE COUNT      73 a      158 c      141 g      78 t
ORIGIN
alignment_scores:
Quality:      692.00      Length:      149
Ratio:        4.908      Gaps:      0

```

Percent Similarity: 94.631 Percent Identity: 82.550

alignment block:  
US-10-040-803-7 x AW463601 ..

Align seg 1/1 to: AW463601 from: 1 to: 450

```

94  AspaNsLeuAenLysProTyrLeuPheserValLeuLeuGlyAlaTPgI 110
3  GATATATCGACAAACCAACCACTTCTCTCTGCTGCTGGAGCCTGCA 52
110 nLeuGlyAsnProGlySerArgSerGlnLysValGlyValAlaLarpyaIg 127
53 GCTGGGAGACCTTGCCCAAGGTCCTCCAGAGGGTGATGCGCTGGGCAC 102
127 LuProHISProValTyrSerTrpGlyGlnGlyAlaCysAlaAspIleAla 143
103 AGCCCCACCTGTGTACTCTCGGAAGAGGGCTCCGCGTCGACATGCGC 152
144 LeuValArgLeuGluArgSerIleGlnPheserGluArgValLeuProI1 160
153 CTGGTGCGCTGGAGCGCGCATCATCTCTGAGGCGGCTCGCCAT 202
160 eCySLeuProAspAlaSerIleHisLeuProProAsnThHisCysTrpI 177
203 CTGCTGCCCCGACTCCACGCTCCAGCTCTCCGAGACCAACATGCTGGA 252
177 IeSerGlyTrpGlySerIleGlnAspGlyValProLeuProHISProGln 193
253 TTGCGCGGCTGGGGAGGCTCCAGATGAGATGCCCCGTGCCACCTCAG 302
194 ThrLeuGlnLysLeuLysValProIleIleAspSerGluValCysSerH1 210
303 ACCCTCAGAGCTGAAGGTCCTCATCATGACTGACACCTGACGCGC 352
210 sleuYrTrpArgGlyAlaGlyGlnGlyProIleHnGlnAspMetLeu 227
353 CCGTACTGCGGGGAGCGCGGAGGCGGCATACCGAGGACATGCTGT 402
227 ySaLaGlyTyrLeuGlnGlyGluArgAspAlaCysLeuGlyAspSer 242
403 GNGCTGGTACCTGAGGGGAGCGGACGCGACCGCTGTGGGCGATTC 449

seq_name: gb_est1:AW383315

seq_documentation_block:
LOCUS      AW383315          507 bp      mRNA      linear      EST 04-FEB-2000
DEFINITION PM1-HT0345-161199-002-h07 HT0345 Homo sapiens cDNA, mRNA sequence.
ACCESSION  AW383315
VERSION    AW383315.1 GI:6887974
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 507)
AUTHORS   HCCP http://www.ludwig.org.br/ORESTES.
TITLE      The FAPESP/LICR Human Cancer Genome Project
JOURNAL    Unpublished (1999)
COMMENT    Contact: Stimpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-27049922
            Fax: +55-11-2707001
            Email: astimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=PM1&f2=PM1-HT0345-161199-002-h07&f3=1999-11-16&f4=1)
            Seq primer: puc 18 forward
            High quality sequence startd 24
```

## FEATURES High quality sequence stop: 507.

Location/Qualifiers  
1..507

/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="HT0345"

/dev\_stage="Adult"  
/note="Organ: head\_neck; Vector: puc18; Site:1: Sma1;  
Site:2: Sma1; A mini-library was made by cloning products  
derived from ORESTES PCR (U.S. Letters Patent application  
No. 196,716 - Ludwig Institute for Cancer Research)  
profiles into the pUC 18 vector. Reverse transcription of  
tissue mRNA and cDNA amplification were performed under  
low stringency conditions."

BASE COUNT 90 a 171 c 129 g 117 t  
ORIGIN

## Percent scores:

Quality: 657.00 Length: 126  
Ratio: 5.256 Gaps: 0  
Percent Similarity: 99.206 Percent Identity: 97.619

alignment\_block:  
US-10-040-803-7 x AW383315 ..

Align seg 1/1 to: AW383315 from: 1 to: 507

```

95 AaLleuAaLlyProTyTrleuPheserValLeuLeuGlyAlaTrpGlnle 111
|||||
128 AACCTGAACAACCACTACCTGTTCTGTGCTGCTGGGGGC.TGCCACT 176
|||||
111 uGlyAsnProGlySerArgSerGlnLysValGlyAlaAlaTrpValGluP 128
|||||
177 GGGGAACCTGGCTCTCGGTCCCAAGAGTGCGTGTGCTGGTGAGC 226
|||||
128 rOhisProValTyTrSerTrpLysGlnGlyAlaCysAlaAspIleAlaLeu 144
|||||
227 CCCACCTGTGTATTCCTGGAAGAGGCGCTGTGCAGACATTCCTG 276
|||||
145 ValArgLeuGlnArgSerIleGlnPheserGlnArgValLeuProIleCy 161
|||||
277 GTGCGTCTGAGCGCTCCATACAGTTCTCAGAGCGGGTCCGCCATCTG 326
|||||
161 sLeuProAspAlaSerIleHisLeuProProAsnThrHisCysTrpIle 178
|||||
327 CCTACTGATGCTCTATCCACCTCCCAACACCCACTGCTGATCT 376
|||||
178 eRgLYTrpGlySerIleGlnAspGlyValProLeuProHisProGlnThr 194
|||||
377 CAGGCTGGGGGAGCATTCCAAAGATGAGTTCCTGCCACCCCTCAGACC 426
|||||
195 LeuGlnLysLeuLysValProIleIleAspSerGlnValCysSerHisLe 211
|||||
427 CTGCAGAAAGCTGAAGGTTCTATCATCATCAGTCGGAAGTCTCAGACCACT 476
|||||
211 uTyTrpArgGlyAlaGlyGlnGlyPro 220
|||||
477 GTACTGGGGGAGATTGAACAGAGGAGCC 504
|||||

```

seq\_name: gb\_est1:AI940071

seq\_documentation\_block:

LOCUS AI940071 330 bp mRNA linear EST 03-AUG-1999  
DEFINITION IL2-CT0031-290799-001-A12 CT0031 Homo sapiens cDNA, mRNA sequence.  
ACCESSION AI940071  
VERSION AI940071.1 GI:5687052  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE 1 (bases 1 to 330)

AUTHORS HCGP <http://www.ludwig.org.br/ORESTES>.  
TITLE The FAPESP/LICR Human Cancer Genome Project  
JOURNAL Unpublished (1999)  
COMMENT Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: [asimpson@ludwig.org.br](mailto:asimpson@ludwig.org.br)  
This sequence was derived from the FAPESP/LICR Human Cancer Genome  
Project. This entry can be seen in the following URL  
(<http://www.ludwig.org.br/scripts/gethtml2.pl?cl=IL2&cl2=IL2-CT0031-290799-001-A12&cl3=1999-07-29&cl4=1>)  
Seq primer: puc 18 forward  
High quality sequence stop: 330.

## FEATURES

source

1..330  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="CT0031"  
/dev\_stage="Adult"

/note="Organ: colon; Vector: puc18; Site:1: Sma1; Site:2:  
Sma1; A mini-library was made by cloning products derived  
from ORESTES PCR (U.S. Letters Patent application No. 196  
716 - Ludwig Institute for Cancer Research) profiles  
into the pUC 18 vector. Reverse transcription of tissue  
mRNA and cDNA amplification were performed under low  
stringency conditions."

BASE COUNT 69 a 101 c 101 g 59 t  
ORIGIN

## alignment\_scores:

Quality: 600.00 Length: 109  
Ratio: 5.505 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-10-040-803-7 x AI940071/rev ..

Align seg 1/1 to reverse of: AI940071 from: 1 to: 330

```

53 GLyLunspSerThrAspSerGluTrpProTrpIleValSerIleGlnly 69
|||||
329 GGCAGAGACAGACACTGACAGAGTGGCCCTGGATCTGAGCATCCAGAA 280
|||||
69 sAsnGlyThrHisCysAlaGlySerLeuLeuThrSerArgTrpValI 86
|||||
279 GAATGGAGACCCACCTGCGCAGGTTCTGTGCTCACCAGCCGCTGGGTGA 230
|||||
86 lEthrAlaAlaHisCysPheLysAspAsnLeuAsnLysProTyRleuPhe 102
|||||
229 TCACCTGCTGCCCACTGTTTCAAGACAACTGAACAACATCATCGTTTC 180
|||||
103 SerValLeuLeuGlyAlaTrpGlnLeuGlyAsnProGlySerArgSerG 119
|||||
179 TCTGTGCTGTGGGGGCTGTGGCAGCTGGGGAAACCTGCTGTGTCCTCA 130
|||||
119 nLysValGlyValAlaAlaTrpValGluProHisProValTyTrpLysG 136
|||||
129 GAAGGTGGGTGTGGCTGGGTGGAGCCCACTGTGTATCTCTGGAAAG 80
|||||
136 LuGlyAlaCysAlaAspIleAlaLeuValArgLeuGlnArgSerIleGln 152
|||||
79 AAGGTGCTGTGCAGACATTCCTGCTGCTGCTCGACGCTCCATACAG 30
|||||
153 PheserGlnArgValLeuProIleCys 161
|||||
29 TTCTCAAGCGGGTCTCGCCCATCTGC 3
|||||

```

seq\_name: gb\_est2:BM389391

```

seq_documentation_block:
LOCUS      BM389391                      703 bp    mRNA    linear    EST 17-JAN-2002
DEFINITION UI-R-CNI-cjK-1-03-0-UI.s1 UI-R-CNI Ratius norvegicus cDNA clone
VERSION    BM389391
KEYWORDS   BM389391.1 GI:18189444
SOURCE     Norway rat.
ORGANISM   Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus
            1 (bases 1 to 703)
REFERENCE   Bonaldo,M.F., Lennon,G. and Soares,M.B.
AUTHORS    Normalization and subtraction: two approaches to facilitate gene
TITLE       discovery
JOURNAL     Genome Res. 6 (9), 791-806 (1996)
MEDLINE    97044477
CONTACT    Soares, MB
PROGRAM    Program for Rat Gene Discovery and Mapping
            University of Iowa
            451 Eckstein Medical Research Building Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: msoares@blue.weeg.uiowa.edu
            The sequence contained an oligo-dT track that was present in the
            oligonucleotide that was used to prime the synthesis of first
            strand cDNA and therefore this may represent a bonafide poly A
            tail. The sequence tag present in the cDNA between the NotI site
            and the oligo-dT track served to identify it as a clone from the
            normalized cervix library cDNA library Preparation: M.B. Soares Lab
            Clone distribution: clones will be available through Research
            Genetics (www.resgen.com) The following repetitive elements were
            found in this cDNA sequence: 1-36, >AT-richLow-complexity
            Seq primer: M13 Forward
            POLY-A=Yes.
FEATURES   Location/Qualifiers
            1..703
            /organism="Rattus norvegicus"
            /strain="Sprague-Dawley"
            /db_xref="taxon:10116"
            /clone="UI-R-CNI-cjK-1-03-0-UI"
            /clone_1ib="UI-R-CNI"
            /dev_stage="adult"
            /lab_host="DH10B (Life Technologies)"
            /note="vector: pT7T3D-Pac (Pharmacia) with a modified
            polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-CNI
            library is a subtracted library derived from the following
            pool of seven normalized rat libraries: normalized rat
            seminal vesicles, normalized rat penis, normalized rat
            bladder, normalized rat cervix, normalized rat brown
            adipose, normalized rat fundus, and normalized rat
            salivary gland. It was constructed according to the
            procedure described by Bonaldo, Lennon & Soares (Genome
            Research Genome 6: 791-806, 1996). For construction of
            the CNI library, plasmid DNA from the pool of normalized
            libraries was electroporated into competent bacteria for
            the production of single-stranded circular DNA. This was
            then used as a tracer in a subtractive hybridization with
            a driver (PCR amplified inserts from a plasmid DNA template
            preparation) comprising: a) a pool of about 34,000 clones
            from the Rat Unigene Set corresponding to plates R-5-AA-NN
            excluding plates R-5-MM and MN. This pool represented 40%
            of the final driver population. b) a pool of about 29,000
            clones from subtracted libraries CA0 and CA1 corresponding
            to plates R-CA0-AWY through R-CA0-AXS, R-CA0-AZX through
            R-CA0-BAZ, R-CA0-BFE through R-CA0-BHJ, R-CA0-BOS,
            R-CA0-BKE, R-CA0-BKG-H, R-CA0-BKJ-K, R-CA0-BKP through
            R-CA0-BKS, R-CA0-BKU-V, R-CA0-BLY through R-CA0-BMA,
            R-CA0-BMC through R-CA0-BME, R-CA0-BNS, R-CA0-BOB through
            R-CA0-BOT, R-CA0-BPA through R-CA0-BPG, R-CA1-BBA through
            R-CA1-BDA, R-CA1-BHZ through R-CA1-BJF, R-CA1-BJR,

```

```

R-CA1-BJT through R-CA1-BKB, R-CA1-BKD, R-CA1-BKF,
R-CA1-BKI, R-CA1-BKT, R-CA1-BLV, R-CA1-BLH through
R-CA1-BLN, R-CA1-BLS, R-CA1-BLU-V, R-CA1-BNR, and
R-CA1-BLE. The resulting pool represented 20% of the
final driver population. c) a pool of about 15,000 clones
from non-normalized libraries CS0, CT0, CU0, CW0, and CX0
and normalized libraries CS0, CT0, CU0, CW0, and CX0
corresponding to plates R-CS0s-CBD through R-CS0s-CBO,
R-CT0s-CAM through R-CT0s-CAX, R-CU0s-CBP through
R-CU0s-CCA, R-CW0s-CCB through R-CW0s-CCM, R-CX0s-CCN
through R-CX0s-CCX, R-CS0-BSD, R-CS0-BTD through R-CS0-BTV
, R-CS0-BVM, R-CT0-BTW through R-CT0-BUP, R-CT0-BVN,
R-CU0-BUQ through R-CU0-BVL, R-CW0-BVV through R-CW0-BMP,
R-CW0-BXN through R-CW0-BXO, R-CX0-BMQ through R-CX0-BXM.
The resulting pool represented 5% of the final driver
population. d) a pool of about 5,000 clones (1,000 from
non-normalized eye library CV0 and 4,000 from normalized
eye library CV1) corresponding to plates R-CV0-BRH through
R-CV0-BRR, R-CV1-BRS through R-CV1-BSC, R-CV1-BSE through
R-CV1-BTC, and R-CV1-BVO through R-CV1-BVO. This pool
represented about 5% of the final driver population. e) A
pool of about 10,000 clones from subtracted library BS2,
BV0 and BV0P (7-9.5 kb cDNA library fraction from rat
whole embryo), and BX0 (0.5-7kb cDNA library fraction from
rat whole embryo) corresponding to plates R-BS2-BDB
through R-BS2-BFB, R-BV0-ANK through R-BV0-ANR, R-BV0P-AOI
through R-BV0P-AOX, and R-BX0-AGY through R-BX0-ASH. The
resulting pool represented 5% of the final driver
population. f) a pool of about 7,000 clones from the
seven non-normalized libraries that make up the tracer
including CY0, CZ0, DA0, DB0, DC0, DD0, and DE0
corresponding to plates R-CY0-BXP through R-CT0-BXZ,
R-CZ0-BYA through R-CZ0-BYL, R-CZ0-BZB-C, R-DA0-BYJ
through R-DA0-BYP, R-DA0-BZD through R-DA0-BZH, R-DB0-BYQ
through R-DB0-BZA, R-DC0-BZI through R-DC0-BZO, R-DE0-CAY
through R-DE0-CBA, R-DD0-BZR through R-DD0-CAL,
R-DD0-CBB-C, and R-DE0-CAB through R-DD0-CAL. The
resulting pool represented about 10% of the final driver
population. g) a pool of about 2,000 clones from the pool
of normalized libraries, CN0, that makes up the tracer.
The corresponding plates are R-CN0-BKW through R-CN0-BLD,
R-CN0-BLG, R-CN0-BLP through R-CN0-BLP, R-CN0-BLF,
R-CN0-BLM-X, R-CN0-BMB, and R-CN0-BMF through R-CN0-BML.
This pool represented 5% of the final driver population.
h) a pool of the 28 most abundant clones in the CN0 pool
corresponding to the following addresses: bkw-a-09-0-UI,
bkw-b-09-0-UI, bkw-b-11-0-UI, bkw-b-10-0-UI, bkw-d-01-0-UI,
bkw-d-06-0-UI, bkw-g-08-0-UI, bkw-h-12-0-UI,
bkw-a-05-0-UI, bkw-a-06-0-UI, bkw-a-11-0-UI, bkw-c-06-0-UI,
bkw-c-09-0-UI, bkw-d-10-0-UI, bkw-a-01-0-UI,
bkw-a-02-0-UI, bkw-f-04-0-UI, bkw-g-07-0-UI, bkw-g-12-0-UI,
bkw-e-12-0-UI, bkw-f-02-0-UI, bkw-a-11-0-UI,
bkw-e-95-0-UI, bkw-d-08-0-UI, bkw-f-02-0-UI, bkw-h-04-0-UI,
bkw-a-05-0-UI, bkw-f-08-0-UI. This pool represented 5%
of the final driver population. i) One abundant CN0 clone
(corresponding to the address bkw-a-11-0-UI) was digested
with Not I and Eco RI and the resulting insert was gel
purified. This purified insert was added directly to the
driver so that it represented 5% of the final driver
population.
TAG_L1B=UI-R-CNI
TAG_L1S=SEU=cervix
TAG_SEO=GACCA*

```

```

BASE COUNT      168 a      178 g      155 t      1 others
ORIGIN

```

```

alignment_scores:
Quality:      586.00      Length:      133
Ratio:        4.883      Gaps:      0
Percent Similarity: 90.226      Percent Identity: 82.707
alignment_block:

```



US-10-040-803-7 x BM389391/rev ..

Align seg 1/1 to reverse of: BM389391 from: 1 to: 703

```

170 PROPANATrHhHsCTrPleSerIleGlySerIleGlnAspG 186
|||||.....:|||||.....:|||||.....:|||||.....:
702 CTTCCACACCACTGCTGATTCGCCGCTGGGAGAACATCCAGGATG 653
186 yValProLeuProHisProGlnThrLeuGlnLeuLysValProIle 203
|||||.....:|||||.....:|||||.....:|||||.....:
652 AGTGCCTCCGACCCCTCAGACCCCTCAGACCTCAGAGCT.CCCATCA 604
203 leAspSerGluValCysSerHisLeuValTrpPargGlyAlaGlyGln 219
|||||.....:|||||.....:|||||.....:|||||.....:
603 TCGACCCCTGAACTCTGCAAAAGTTGTACTGCGCGGAGCTGTCAGGAA 554
220 ProIleThrGluAspMetLeuCysAlaGlyTrpLeuGlnGlyGluArg 236
|||||.....:|||||.....:|||||.....:|||||.....:
553 GCCATCACCGAGGACATCTGTGTCTGTGTACTCGAAGGAGGAGCGGGA 504
236 PALAcysLeuGlyAspSerGlyGlyProLeuMetCysGlnValAspGly 253
|||||.....:|||||.....:|||||.....:|||||.....:
503 CGCTGTGTGGGCGACTGTGGGGGTCCTGATGTGCCAGGTGATGAC 454
253 latrPLeuLeuAlaGlyIleIleSerTrpGlyGluGlyCysAlaGlyArg 269
|||||.....:|||||.....:|||||.....:|||||.....:
453 ACTGGCTACTGACGGGCAATATCACTGGGAGAGAGGCTGCGCGGAGCG 404
270 AsnArgProGlyValTrpIleSerLeuSerAlaHisArgSerTrpValG 286
|||||.....:|||||.....:|||||.....:|||||.....:
403 AACCGCGCGGCGGTACACGACCTCTAGCTCACGCCCTCGGTGCA 354
286 uLysIleValGlnGlyValGlnLeuArgGlyArgAlaGlnGlyGly 302
|||||.....:|||||.....:|||||.....:|||||.....:
353 GAGGATCTGTGCAAGGGGTGCAAGCTGCGAGGCGCTTGCGGACAGTGG 305

```

seg\_name: gb\_est2:BG480197

seg\_documentation\_block:

LOCUS BG480197 1001 bp mRNA linear EST 21-MAR-2001  
 DEFINITION 602530380F1 NIH\_MGC\_21 Homo sapiens cDNA clone IMAGE:465395 5',  
 mRNA sequence.

ACCESSION BG480197

VERSION BG480197.1 GI:13412476

KEYWORDS EST.

SOURCE human.

ORGANISM

Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 1 (bases 1 to 1001)

AUTHORS

TITLE NIH-MGC http://mgc.nci.nih.gov/

JOURNAL

COMMENT Contact: Robert Strausberg, Ph.D.

Tissue Procurement: ATCC

CDNA Library Preparation: Ling Hong/Rubin Laboratory

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LNL at:

http://image.llnl.gov

plate: LICM1440 row: P column: 20

High quality sequence stop: 536.

Location/Qualifiers

1..1001

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:465395"

/clone\_lib="NIH\_MGC\_21"

/tissue\_type="choriocarcinoma"

/lab\_host="DH10B (phage-resistant)"

/note="Organ: placenta; Vector: pOTB7; Site:1; XhoI;

Site 2: EcoRI; cDNA made by oligo-dT priming.  
 directionally cloned into EcoRI/XhoI sites using the  
 following 5' adaptor: GGCAGAG(C). Size-selected >500bp  
 for average insert size 1.8kb. Library constructed by  
 Ling Hong in the Laboratory of Gerald M. Rubin (University  
 of California, Berkeley) using ZAP-cDNA synthesis kit  
 (Stratagene) and Superscript II RT (Life Technologies)."

BASE COUNT 254 a 255 c 292 g 200 t  
 ORIGIN

alignment\_scores:

Quality: 580.50 Length: 267

Ratio: 2.977 Gaps: 24

Percent Similarity: 73.034 Percent Identity: 64.045

alignment\_block:

US-10-040-803-7 x BG480197 ..

Align seg 1/1 to: BG480197 from: 1 to: 1001

```

1 MetValValSerGlyAlaProProAla.LeuGlyGlyGlyCysLeu.Gly 16
|||||.....:|||||.....:|||||.....:|||||.....:
53 ATGGGTCTTCTGTGAGGCCACCCAGACCTGGGTGGGCTGTCTGGC 102
17 ThrPheThrSerLeuLeuLeuAla.SerThrAlaIleLeuAsnAla. 32
|||||.....:|||||.....:|||||.....:|||||.....:
103 ACCCTTCACCTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 152
33 AlaArgIleProValProProAlaCysGlyLysProGlnLeuAsnAr 49
|||||.....:|||||.....:|||||.....:|||||.....:
153 GCCAGATACCTGTTCCGCCAGCCTGTGGGAGGCCAGCAGCCTGAAC 202
49 gValValGlyGly.GluAspSerThrAspSerGluTrpProTrpIleVal 65
|||||.....:|||||.....:|||||.....:|||||.....:
203 GGTGTGGGGGCTGAGGAGCAGCAGCAGCAGGAGTGTGATCTGTG 252
66 SerIleGlnLysAsn.GlyThrHisHisCysAla.GlySerLeu.Leu 81
|||||.....:|||||.....:|||||.....:|||||.....:
253 AGCTTCAGAGAGATAGGAGGCCACACCTGCGCAGATTTCTAGCTCAC 302
81 rSerArgTrpValIleThrAlaAlaHisCysPheLysAspAsnLeuAsn 98
|||||.....:|||||.....:|||||.....:|||||.....:
303 CAGACGCTGGGTGATTCACCTGCTGCCCTGTTCAAGACCAACTGACA 352
98 ySProTyrLeu.PheSerValIleLeuGlyAlaTrp...GlnLeuGlyAs 113
|||||.....:|||||.....:|||||.....:|||||.....:
353 AAACATACCTGTTACTCTGTGCTGCTAGAGGGGCTGGCAAGCTGGGAA 402
113 nProGly...SerArgSerGlnLysValGlyValAlaIleTrp.ValGlu.Pr 128
|||||.....:|||||.....:|||||.....:|||||.....:
403 CCCATGCTCTCGAGTCCCAAGAGAGGTGGTGTTCCTGGGATGGAGACC 452
128 OHsPProValTrpSerTrpLysGlnGlyAlaCysAlaAsp...IleAla 144
|||||.....:|||||.....:|||||.....:|||||.....:
453 CCACCTGTATTCCTGAGCGAGAGGTGCTGTGCGAGCAATTTGTTCTG 502
144 euValArgLeuGluArgSerIleGln.PheSerGlu.ArgValLeuPro 160
|||||.....:|||||.....:|||||.....:|||||.....:
503 GATCAGCTCTGAGCGCTCATCAAGTTCTCAGAGCGGGTACTGACCA 552
160 leCysLeuProAspAla.....SerIleHisLeuProProAsnThrHis 174
|||||.....:|||||.....:|||||.....:|||||.....:
553 TCGGCTACTGATGCTCTTCATCCACTCAGACGTCACAAAGAACGCAAT 602
175 CysTrpIleSerGlyTrp...GlySerIleGlnAspGly.ValProLeu. 189
|||||.....:|||||.....:|||||.....:|||||.....:
603 GAGTGGATCTCAGAGCTAGAGGAGGCCATCCAAAGATGGAATCACTTG 652
190 ..ProHisProGlnThrLeuGlnLysLeu...LysValPro.IleIleAs 204
|||||.....:|||||.....:|||||.....:|||||.....:
653 CACCCAACTCAGAGGCTGAGAAAGCTAGAAAGGTTCTTCACTATCGA 702

```

```
204 pSerGluValCysSerHisLeuTyrTrpArgGlyAlaGly.GlnGlyPro 220
    |||  ::  ::  ::  |||  ::  |||  |||
703 CTCGGCAGTCCGACAGACATCTGTACTTGGCGGAGGACAGACAGACCA 752
    |||  ::  ::  |||  ::  |||  |||
221 IleHisGluAspMet.....LeuCysAlaGlyTyrLeuGluGlyG1 234
    |||  ::  ::  |||  ::  |||  |||
753 GGGACCCAGTTCATCTGTAGACACTGTGTGTGTGTGTGTGTGTGTGTGT 802
    |||  ::  ::  |||  ::  |||  |||
234 uArgAspAlaCysLeuGly 240
    :  |||  ::  |||
803 AGGTAAGCAGGAGATTGGA 821
```

\_\_\_\_\_

```

137 GACATGAGGCTTCTGAGAGTAAGTGGCCCTGCGAGGAGGCTGAGATT 186
69 .....LysAsnGlyThrHisCysAlaGlySerLeuLeuThrSerAr 83
187 AATATTAACTACTGATACATTTCTGTGGAGGCTCTCTATCCACCCACA 236
83 gTTPValIleThrAlaAlaHisCysPheLysAspAsnLeuAsnLysProt 100
237 GTGGGTCTCTACTGCGGACACCTGTGGGACCCGACATATAAAAGCCAC 286
100 yLeuPheSerValLeuLeuGlyAlaTrpLeu.....GlyAsnPro 114
287 AGCTCTTCGGGTGACGCTTCGTGAGCAGTATCTATGAGGAGC... 333
115 GlySerArgSerGlnLysValGlyAlaTrpValGluTrpHisProVa 131
334 .....CAGCTCTCTCTTTTAAACCGGATCGTGGTACCCCA 371
131 lTySerTrpLysGluGlyAlaCysAlaAspIleAlaLeuValArgLeu 148
372 CTATTACACGCGCGAGGCTGG...GCAGACGTTCGCCCTCTGAGCTTG 418
148 lueArgSerIleGlnPheSerGluArgValLeuProIleCysLeuProAsp 164
419 AGCTCCCTGTGAATGTCTCCACCATATCCACCCATATCCCTGCCCCCT 468
165 AlaSerIleHisLeuProProAsnThrHisCysTrpIleSerGlyTrpAl 181
469 GCCTCGGAGACCTTCCCTCGGAGATCGTGTGGGTACACAGCGGGG 518
181 ySerIleGlnAspGlyValProLeuProHisProGlnThrLeuGlnLysL 198
519 CGACATTGATATGACGAGGCTCTCCACCTCTTATCTCTGAAAGCAAG 568
198 euLysValProIleIleAspSerGluValCysSerHisLeuTrpArg 214
569 TGAAGGTTCCTCATTTGGAAACAGCTGTGACCGGAGTACCACT 618
215 GlyAlaGlyGlnGly.....ProIleThrGluAsp...MetLeuCy 227
619 GGCCTCTACACGAGATGATTTTCCCATTTGTCATGATGATGCTGTG 668
227 salaglyTyLeuGluGlyGluArgAspAlaCysLeuGlyAspSerGlyG 244
669 TGCTGA.....AATACAGAGAGACCTCTGCCAGAGGCGATTCAGGGG 712
244 lYProLeuMetCysGlnValAlaAspGlyAlaTrpLeuLeuAlaGlyIle 260
713 GGCCTGTGTGTGCAAGTGAAGGATGCTGCTGACGAGCAGAGATGTGTC 762
261 SerTPGlyGluGlyCysAlaGluArgAsnArgProGlyValTyTrile 277
763 ACCTGGGGGAGGCTGCGCACAGCCCAACAGCTGCGATCTACACCCG 812
277 rLeuSerAlaHisArgSerTrpValGluLysLeuVal 289
813 GGTGACATACCTACTTACACTGATCCACCGCATATGTC 849
seq_name: /cgn2_6/plodata/1/ina/5B_COMB.seq:US-08-978-404B-20
seq_documentation_block:
; Sequence 20, Application US/08978404B
; Patent No. 5968782
; GENERAL INFORMATION:
; APPLICANT: Stevens, Richard L.
; TITLE OF INVENTION: MAST CELL PROTEASE THAT CLEAVES
; NUMBER OF SEQUENCES: 74
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wolf, Greenfield & Sacks, P.C.
; STREET: 600 Atlantic Avenue
; CITY: Boston

```

```

STATE: MA
COUNTRY: U.S.A.
ZIP: 02210-2211
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/978,404B
FILING DATE: 25-NOV-97
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/032,354
FILING DATE: 04-DEC-1996
ATTORNEY/AGENT INFORMATION:
NAME: Plumer, Elizabeth R.
REGISTRATION NUMBER: 36,637
REFERENCE/DOCKET NUMBER: B0801/7090
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-720-3500
TELEFAX: 617-720-2441
TELEX:
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 1108 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-978-404B-20

```

```

alignment_scores:
Quality: 554.00 Length: 279
Ratio: 2.916 Gaps: 9
Percent Similarity: 68.100 Percent Identity: 41.577

```

```

alignment_block:
US-10-040-803-7 x US-08-978-404B-20

```

```

Align seg 1/1 to: US-08-978-404B-20 from: 1 to: 1108

```

```

21 LeuLeuLeuLeuAlaSerThrAlaIleLeuAsnAlaAlaArgIleProVa 37
49 CTGCTGTGTGTGTCGACATGTCCTCTGCGTGTGTGTGTGTGTGTGTGT 98
37 lPro...ProAlaCysGlyLysProGlnGlnLeuAsnArgValGlyG 53
99 CCTCGCCCGCCAGC.....AATCAGCAGAGTGGCGCATGCTGGGAG 136
53 lYgluAspSerThrAspSerGluTrpProThrIleValSerIleGln... 68
137 GACATGAGGCTTCTGAGAGTAAGTGGCCCTGCGAGTGTGAGCTGAGATT 186
69 .....LysAsnGlyThrHisCysAlaGlySerLeuLeuThrSerAr 83
187 AATTTAACTACTGATACATTTCTGTGGAGGCTCTCTATCCACCCACA 236
83 gTTPValIleThrAlaAlaHisCysPheLysAspAsnLeuAsnLysProt 100
237 GTGGGTCTCTACTGCGGACACCTGTGGGACCCGACATATAAAAGCCAC 286
100 yLeuPheSerValLeuLeuGlyAlaTrpLeu.....GlyAsnPro 114
287 AGCTCTTCGGGTGACGCTTCGTGAGCAGTATCTATGAGGAGC... 333
115 GlySerArgSerGlnLysValGlyAlaTrpValGluTrpHisProVa 131
334 .....CAGCTCTCTCTTTTAAACCGGATCGTGGTACCCCA 371
131 lTySerTrpLysGluGlyAlaCysAlaAspIleAlaLeuValArgLeu 148
372 CTATTACAGCGCGAGGCTGG...GCAGACGTTCGCCCTCTGAGCTTG 418

```

```

148 luargserileglnpheserleuargvalleuproileucysleuproasp 164
149 AGGTCCTGTAATGTCTCCACCATATCCACCCATATCCCGCCCT 468
165 AlaserlehisleuproasnrthhsCysTrpIleSerGlyTrpI 181
166 GCTTCGGAGACCTTCCCGCTGGGACATCTGCTGGGTGACAGCGTGGG 518
181 yserileglnaspllyalproleuprohisproglnthleuglnysl 198
182 CGATTTGATATATGACGAGCTCTCCACCTCTTATCTCTTAAGCAAG 568
198 eulysvalproilleaspsersleuValCysSerhisleuTrpArg 214
199 TGAGGTCCCATTTGTGTGAAACAGCCTGTGTGACCGGAGTACCACT 618
215 glYalaGlyGlnGly.....ProIleThrGlnAsp...MetLeuCy 227
219 GGCTCTACAGCGGAGATGATTTCCATGTCTCATGATGCGATGCTGTG 668
227 salaellyrleuGlnGlyluargaspAlaCysLeuGlyAspserslyg 244
228 TTTTGA.....AATACAGAGAGAGACTCTCTGCGAGCGGATTCAGGG 712
244 lYProleuMetCysGlnValaspllyaltrpleuleuAlaGlyIle 260
245 GGCACCTGTGTGCAAGTGAAGGTACTGCTGCGAGCGAGAGTGTG 762
261 SerTrpGlyGlnGlyCysAlaGlnArgasnArgProGlyValTrpIle 277
262 ACTGTGGGGAGGGGCTGCGACAGCCCAAGCTGTCATCTACACCG 812
277 rleuserAlaHisArgSerTrpValGlnLysIleVal 289
280 GGTGACATCTACTTATGACTGATGATCCACCGCTATGTC 849
813 GGTGACATCTACTTATGACTGATGATCCACCGCTATGTC 849

seq_name: /cgn2_6/p1odata/1/ina/5B_COMB.seq:US-08-978-404B-4

seq_documentation_block:
; Sequence 4, Application US/08978404B
; Patent No. 5968782
; GENERAL INFORMATION:
; APPLICANT: Stevens, Richard L.
; TITLE OF INVENTION: MAST CELL PROTEASE THAT CLEAVES
; NUMBER OF SEQUENCES: 74
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wolf, Greenfield & Sacks, P.C.
; STREET: 600 Atlantic Avenue
; CITY: Boston
; STATE: MA
; COUNTRY: U.S.A.
; ZIP: 02210-2211
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/978,404B
; FILING DATE: 25-NOV-97
; CLASSIFICATION: 435
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: 60/032,354
; FILING DATE: 04-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Plumer, Elizabeth R.
; REGISTRATION NUMBER: 36,637
; REFERENCE/DOCKET NUMBER: B0801/7090
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-720-3500
; TELEFAX: 617-720-2441

```

```

;
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1097 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-978-404B-4

alignment_scores:
    Quality: 538.50      Length: 288
    Ratio: 2.849        Gaps: 10
    Percent Similarity: 65.625    Percent Identity: 40.625

alignment_block:
US-10-040-803-7 x US-08-978-404B-4 ..

Align seg 1/1 to: US-08-978-404B-4 from: 1 to: 1097

21 leuleuleuAlaserThr.....AlalleuAsnAlaAlaAr 34
22 TTTTGA.....AATACAGAGAGAGACTCTCTGCGAGCGGATTCAGGG 712
38 CTGCTGCTGCTGCGACTGTCTCCCGCTGCTAGTGTGACCGCGGCC 87
34 gileProvalProProAlaCysGlyLysProGlnGlnLeuAsnArgVal 51
88 TTGCCCATCTCAAGCAGCGAGTGGG.....ATTG 116
51 alGlyGlyGlnAspsersThrAspsersGlnTrpProTrpIleValSer 67
117 TGGGAGGAGCAGAGGCTTCTGAAGTAAAGTGGCTGCGAGGTAGCTG 166
68 GlnLysAsnGlyThr.....HisHisCysAlaGlySerLeuThr 81
167 AGATTAAATTCAGCTTCTGATGATGATTTCTGTGGCGGCTCCCTATCA 216
81 rserArgTrpValIleThrAlaAlaHisCysPheLysAsnLeuAsn 98
217 CCACAGTGGGTCTCAGTGGCGCACACTGTGTGGAGTGCATCAAA 266
98 yAsProTrpLeuPheSerValLeuLeuGlyAlaTrpGlnLeuGlyAsnPro 114
267 GCCCAGAGCTTCTCCGTGACAGCTT..... 292
115 GlySerAlaSerGln.....LysValGln 122
293 .....CGTACACAGTATCTATGATGCGACAGTACTGATGAA 336
122 yAlaAlaTrpValGlnProHisProValTrpSerTrpLysGlnGlyAla 139
337 CCGAGCGGTTGTGACACCCCGAC...TACTACACAGTGCAGGATGGG... 379
139 yAlaAspIleAlaLeuValArgLeuGlnArgSerIleGlnPheSerGln 155
380 ..GCAGCATTTGCTGCTGCGAGCTTGAAGATCCCTGTAATGTCTCCACC 427
428 CATATCCACCCCATATCCCTGCCCTGCGAGACCTTCCCTCGGG 477
172 nThrHisCysTrpIleSerGlyTrpGlySerIleGlnAspGlyValPro 189
478 GACTTCTGCTGGTAAACAGGCTGGGCGCACATCTATATGACGAGCTC 527
189 eProHisProGlnTrpLeuGlnLysLeuValProIleLeuAspSer 205
528 TCCGCGACCTTATCTCTGAAGCAAGTGAAGTCCCATTTGTGAAAC 577
206 GlnValCysSerHisLeuTrpArgGlyAlaGlyGlnGly..... 219
578 AGCTGTGTGATCGAGATACACACTGCGCTCTACACAGGAGATGAT 627
220 .ProIleThrGlnAsp...MetLeuCysAlaGlyTrpLeuGlnGlyGln 235

```



\_\_\_\_\_

1000

```

; Patent No. 5955431
; GENERAL INFORMATION:
; APPLICANT: Stevens, Richard L.
; APPLICANT: Huang, Chifu
; TITLE OF INVENTION: MAST CELL PROTEASE PEPTIDE
; TITLE OF INVENTION: INHIBITORS
; NUMBER OF SEQUENCES: 65
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wolf, Greenfield & Sacks, P.C.
; STREET: 600 Atlantic Avenue
; CITY: Boston
; STATE: MA
; COUNTRY: U.S.A.
; ZIP: 02210-2211
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/016,366A
; FILING DATE: January 30, 1998
; CLASSIFICATION: 530
; PRIORITY DATA:
; APPLICATION NUMBER: 60/037,090
; FILING DATE: 05-FEB-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Plumer, Elizabeth R.
; REGISTRATION NUMBER: 36,637
; REFERENCE/DOCKET NUMBER: B0801/7093
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-720-3500
; TELEFAX: 617-720-2441
; TELEX:
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1128 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-09-016-366A-20

alignment_scores:
      Quality: 530.50      Length: 288
      Ratio: 2.837      Gaps: 9
      Percent Similarity: 64.931      Percent Identity: 39.236

      ment block:
      0-040-803-7 x US-09-016-366A-20 ..

Align seg 1/1 to: US-09-016-366A-20 from: 1 to: 1128
20 SerLeuLeuLeuLeuAlaSerThrAlaIleuAsnAlaAlaArgIlePr 36
   ::::::::::::::::::::
5 AATGTGCTGTCTGTGGCGCTGCTCTG.....GCGAGCCGGCGCTA 48
   ::::::::::::::::::::
36 oValProProAlaCysGlyLysProGlnGlnLeuAsnArgValAlaGly 53
   ::::::::::::::::::::
49 CGCGGCCCTGTGCGCCAGCGCCCTGCAGCGAGTGGCGATCGTGGGG 98
   ::::::::::::::::::::
53 LysAlaSerThrAspSerGlyLysProGlnGlnLeuAsnArgValAlaGly 69
   ::::::::::::::::::::
99 GTCAAGAGGCCCCCGAGAGAGTGGCGCTGCGAGTGGAGCTGAGATC 148
   ::::::::::::::::::::
70 AsnGly.....ThrHisCysAlaGlySerLeuLeuThrSerAr 83
   ::::::::::::::::::::
149 CACGGCCCATGATGATGATGATGATGATGATGATGATGATGATGATG 198
   ::::::::::::::::::::
83 GTTValIleThrAlaAlaHisCys.....PheLysAsp.... 94
   ::::::::::::::::::::
199 GTGGGTGCTGACCGACGCGACTGCTGTGGACCGAGCTCAAGATCTGG 248

```

```

95 .....AsnLeuAsnLysProTyrLeuPheSerValLeuLeu 106
   ::::::::::::::::::::
249 CCGCCCTCAGAGGTGCAACTGCGGGAGACGACCTCTAC..... 286
   ::::::::::::::::::::
107 GAlaATrPglInLeuGlnLysProGlySerArgSerGlnLysValAla 123
   ::::::::::::::::::::
287 .....TACAGAGACCACTGCTGCGCGT 309
   ::::::::::::::::::::
123 lAlaTrValIGluProHisPro...ValTyrSerTrpLysGlnGlyAla 139
   ::::::::::::::::::::
310 CACGAGATCATGCTGCTACACCGATTTACACCGCCGATCGA.... 355
   ::::::::::::::::::::
139 yAlaAspIleAlaLeuValArgLeuGluArgSerIleGlnPheSerGlu 155
   ::::::::::::::::::::
356 ..GCGGACATCGCCCTGTGAGCTGAGAGACCGGTGAAGTCTCCAGC 403
   ::::::::::::::::::::
156 ArgValLeuProIleCysLeuProAspAlaSerIleHisLeuProHis 172
   ::::::::::::::::::::
404 CAGTCCACAGGTGACCTGCTGCGCCCTGCTCAGAGACCTTCCCGGG 453
   ::::::::::::::::::::
172 nThrHisCysTrpIleSerGlyTrpGlySerIleGlnAspGlyValPro 189
   ::::::::::::::::::::
454 GATGCGGTGCTGCTGACAGCTGCGGGCGATGAGATGATGAGCGCC 503
   ::::::::::::::::::::
189 euProHisProGlnThrLeuGlnLysLeuLysValProIleLeuAspSer 205
   ::::::::::::::::::::
504 TCCACCGCCGATTTCTCTGTAAGCAGTGAAGTCCCATATGTAAGAAC 553
   ::::::::::::::::::::
206 GluValCysSerHisLeuTyrTrpArgGlyAlaGlnGlnGlyPro.... 220
   ::::::::::::::::::::
554 CACATTTGTGACGCCAAATPACACCTTGCGCTTACAGGAGACGACGT 603
   ::::::::::::::::::::
221 .....lIeThrGluAspMetLeuCysAlaGlyTyrLeuGlnGlyAla 235
   ::::::::::::::::::::
604 CGCATGCTGCTGCTGACGACATGCTGTGTGCGGG.....AACCCGGA 647
   ::::::::::::::::::::
235 rGAspAlaCysLeuGlnLysSerGlyLysProLeuMetCysGlnValAsp 251
   ::::::::::::::::::::
648 GGGACTCATGCCAGGCGGCGCTGCGAGAGGCGCCGTGTGTGCAAGTGAAT 697
   ::::::::::::::::::::
252 GAlaATrPglLeuAlaGlyIleIleSerTrpGlyGlnGlyCysAlaG 268
   ::::::::::::::::::::
698 GCACTGCTGCTGACGCGCGGTGTGCTGAGTGGGCGGAGGCTGTGCCA 747
   ::::::::::::::::::::
268 uArgAsnArgProGlyValTyrIleSerLeuSerAlaHisArgSerTrp 285
   ::::::::::::::::::::
748 GCCACCGCGCTGCTGATACACCGCTGTGACCTACTGAGTGA 797
   ::::::::::::::::::::
285 alGluLysIleVal 289
   ::::::::::::::::::::
798 TCCACCACTATGTC 811

seq_name: /cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-978-404B-15
seq_documentation_block:
; Sequence 15, Application US/08978404B
; Patent No. 5968782
; GENERAL INFORMATION:
; APPLICANT: Stevens, Richard L.
; TITLE OF INVENTION: MAST CELL PROTEASE THAT CLEAVES
; NUMBER OF SEQUENCES: 74
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wolf, Greenfield & Sacks, P.C.
; STREET: 600 Atlantic Avenue
; CITY: Boston
; STATE: MA
; COUNTRY: U.S.A.
; ZIP: 02210-2211
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible

```



```
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/978,404B
FILING DATE: 25-NOV-97
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/032,354
FILING DATE: 04-DEC-1996
ATTORNEY/AGENT INFORMATION:
NAME: Plumer, Elizabeth R.
REGISTRATION NUMBER: 36,637
REFERENCE/DOCKET NUMBER: B0801/7090
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-720-3500
TELEFAX: 617-720-2441
TELEX:
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 1128 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-978-404B-15

alignment_scores:
Quality: 530.50 Length: 288
Ratio: 2.837 Gaps: 9
Percent Similarity: 64.931 Percent Identity: 39.236

alignment_block:
US-10-040-803-7 x US-08-978-404B-15 ..

Align seg 1/1 to: US-08-978-404B-15 from: 1 to: 1128

20 SerLeuLeuLeuAlaSerThrAlaLeuAlaAlaArgIlePr 36
   :::::::::::::::::::: :
5 AATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 48
   :::::::::::::::::::: :
36 oValProAlaCysGlyLysProGlnLeuAlaArgValAlaGly 53
   :::::::::::::::::::: :
49 CGGGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 98
   :::::::::::::::::::: :
53 LysLysPheThrAspSerGluTrpTrpIleAlaSerIleGlnLys 69
   :::::::::::::::::::: :
99 GTGAGAGGCGCCGACGAGCAAGTGGCTGCTGCTGCTGCTGCTG 148
   :::::::::::::::::::: :
70 AsnGly.....ThrHisCysAlaGlySerLeuLeuThrSerAr 83
   :::::::::::::::::::: :
149 CAGGCGCCATCTAGATGATGCACTTCTGCGGGGCTCCCTATCCACCCCA 198
   :::::::::::::::::::: :
83 gTrpValIleThrAlaAlaHisCys.....PheLysAsp... 94
   :::::::::::::::::::: :
199 GTGGGTGTGACCGCGACGCGACGCTGGGACCGGACGTCAGATCTGG 248
   :::::::::::::::::::: :
95 .....AsnLeuAsnLysProGlyLeuPheSerValLeuLeu 106
   :::::::::::::::::::: :
249 CCGGCTCAGAGGTGCACTGCGGGAGACGACCTCTAC..... 286
   :::::::::::::::::::: :
107 GlyAlaTrpGlnLeuLysAnProGlySerArgSerGlnLysValGlyAla 123
   :::::::::::::::::::: :
287 .....TACCAGGACCACTGCTGCGCGGT 309
   :::::::::::::::::::: :
123 LAlaTrpValGluProHisPro...ValIleSerTrpLysGluGlyAlaC 139
   :::::::::::::::::::: :
310 CAGCAGAGATCATGTCACCCAGATTCTACACCGCCGATCGGA... 355
   :::::::::::::::::::: :
139 LysAlaSplLeuAlaLeuValArgLeuGluArgSerIleGlnPheSerGlu 155
   :::::::::::::::::::: :
356 ..GGGACATCGCTGCTGAGAGCTGAGAGGCGGCTGTAAGTCTCCACG 403
   :::::::::::::::::::: :
156 ArgValLeuProIleCysLeuProAspAlaSerIleHisLeuProProAs 172
```

```
404 CAGCTCCACAGCGTCACACCTGCCCCCTGCCCTCAGAGACCTCCCGCGG 453
   :::::::::::::::::::: :
172 nThrHisCysTrpIleSerGlyTrpGlySerIleGlnAspGlyValProL 189
   :::::::::::::::::::: :
454 GATCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 503
   :::::::::::::::::::: :
189 euProHisProGlnThrLeuGlnLysLeuLysValProIleIleAspSer 205
   :::::::::::::::::::: :
504 TCCACCGCGCATTTCTCTGTAGAGCATGAGGTGCCCATTAATGAAAC 553
   :::::::::::::::::::: :
206 GluAlaCysSerHisLeuTyTrpArgGlyAlaGlyGlnIlePro..... 220
   :::::::::::::::::::: :
554 CACATTGTGTACGCCAATAACACACCTTGGCGGCTACACGGAGACGACGT 603
   :::::::::::::::::::: :
221 .....IleThrGluAspMetLeuCysAlaGlyTyTrpLeuGluGlyL 235
   :::::::::::::::::::: :
604 CCGCATGCTCCGTGACGACATGCTGTGCGCGG.....AACCCCGGA 647
   :::::::::::::::::::: :
235 rGAspAlaCysLeuGlyAspSerGlyGlyProLeuMetCysGlnValAsp 251
   :::::::::::::::::::: :
648 GGGACTATGCTCCAGGCGGACTCCGAGAGGCGCCCTGTGTGCAAGTGAAT 697
   :::::::::::::::::::: :
252 GlyAlaTrpLeuLeuAlaGlyIleIleSerTrpGlyGluGlyCysAlaG 268
   :::::::::::::::::::: :
698 GGCACTGTGCTGACGGCGGCGGTGTCAGTGGGCGAGGCTGTGCCCA 747
   :::::::::::::::::::: :
268 uArgAsnArgProGlyValTyTrpIleSerLeuSerAlaHisArgSerTrp 285
   :::::::::::::::::::: :
748 GCCACCGCGCTGCTGACATCTACACCGCTGACCTACTTGGACTGGA 797
   :::::::::::::::::::: :
285 aGluLysIleVal 289
   :::::::::::::::::::: :
798 TCCACCACTATGTC 811

seq_name: /cgn2.6/plodata/1/lna/5B_COMB.seq:US-09-016-366A-16

seq_documentation_block:
; Sequence 16, Application US/09016366A
; Patent No. 5955431
; GENERAL INFORMATION:
; APPLICANT: Stevens, Richard L.
; APPLICANT: Huang, Chifu
; TITLE OF INVENTION: MAST CELL PROTEASE PEPTIDE
; TITLE OF INVENTION: INHIBITORS
; NUMBER OF SEQUENCES: 65
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wolf, Greenfield & Sacks, P.C.
; STREET: 600 Atlantic Avenue
; CITY: Boston
; STATE: MA
; COUNTRY: U.S.A.
; ZIP: 02210-2211
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/016,366A
; FILING DATE: January 30, 1998
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/037,090
; FILING DATE: 05-FEB-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Plumer, Elizabeth R.
; REGISTRATION NUMBER: 36,637
; REFERENCE/DOCKET NUMBER: B0801/7093
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-720-3500
; TELEFAX: 617-720-2441
; TELEX:
```



Align seg 1/1 to: US-08-978-404B-11 from: 1 to: 1154

```

       /cgn2_6/prodata/1/lna/5B_COMB.seq;us-09-016-366A-18
Documentation_block:
Sequence 18, Application US/09016366A
Invent No. 5955431
GENERAL INFORMATION:
APPLICANT: Stevens, Richard L.
APPLICANT: Huang, Chifu
TITLE OF INVENTION: MAST CELL PROTEASE PEPTIDE
TITLE OF INVENTION: INHIBITORS
NUMBER OF SEQUENCES: 65
CORRESPONDENCE ADDRESS:
ADDRESSEE: Wolf, Greenfield & Sacks, P.C.
STREET: 600 Atlantic Avenue
City: Boston
STATE: MA
COUNTRY: U.S.A.
Zip: 02210-7211
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/016,366A
FILING DATE: January 30, 1998
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/037,090
FILING DATE: 05-FEB-1997
ATTORNEY/AGENT INFORMATION:
NAME: Plummer, Elizabeth R.
REGISTRATION NUMBER: 36,637
REFERENCE/DOCKET NUMBER: B0801/7093
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-720-3500
TELEFAX: 617-720-2441
TELEX:
FORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 1137 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
-016-366A-18
ment_scores:
Quality: 529.50 Length: 288
Ratio: 2.832 Gaps: 9
Percent Similarity: 64.931 Percent Identity: 39.236
ment_block:
..
seq 1/1 to: US-09-016-366A-18 from: 1 to: 1137
20 SerLeuLeuLeuLeuAlaSerThrAlaIleLeuAsnAlaLArgTllePr 36
:::|||||:::|||||
3 AACTCTCTCTCTGTGGCCCTGCCCCCTGCTG.....GCAGCGCGGCCCTA 46
36 oValProProlAlaCysGlyLysProGlnGlnLeuAsnArgValAlaGlyg 53
:::|||||:::|||||
47 CGGGGCGCCCTGCCCAAGGCACAGGCCCTGCAGCAGATGGGCATCTGCGGG 96
53 LysLaspSerThrAspSerGluTrpProTrpIleValSerIleGlnLys 69
|||||||:::|||||
97 GTGAGAGAGCGCCCAAGGAGCAAGTGCGCCCTGCAGGTCGTGAGTAGCTC 146
0 AsnGlyL.....ThrHisHisCysAlaGlySerLeuLeuThrSerAr 83

```

```

seq_name:/g9n2_6/plodata/1/ina/5B_COMB.seq:US-09-016-366A-18
seq documentation block:
? Sequence 18 Application US/09016366A
? Patent No. 5955421
? GENERAL INFORMATION:
? APPLICANT: Stevens, Richard L.
? APPLICANT: Huang, Chifu
? TITLE OF INVENTION: MAST CELL PROTEASE PEPTIDE
? TITLE OF INVENTION: INHIBITORS
? NUMBER OF SEQUENCES: 65
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: Wolf, Greenfield & Sacks, P.C.
? STREET: 600 Atlantic Avenue
? CITY: Boston
? STATE: MA
? COUNTRY: U.S.A.
? ZIP: 02210-2211
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/016,366A
FILING DATE: January 30, 1998
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/037,090
FILING DATE: 05-FEB-1997
ATTORNEY/AGENT INFORMATION:
NAME: Plummer, Elizabeth R.
REGISTRATION NUMBER: 36,637
REFERENCE/DOCKET NUMBER: B0801/7093
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-720-3500
TELEFAX: 617-720-2441
TELEX:
? INFORMATION FOR SEQ ID NO: 18:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 1137 base pairs
? type: nucleic acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: cDNA
US-09-016-366A-18

Alignment_scores:
Quality: 529.50 Length: 288
Ratio: 2.832 Gaps: 9
Percent Similarity: 64.931 Percent Identity: 39.236

alignment_block:
US-10-040-803-7 x US-09-016-366A-18 ..

Align seg 1/1 to: US-09-016-366A-18 from: 1 to: 1137

20 SerLeuLeuLeuAlaSerThrLalLeuAsnAlaValArgIlePr 36
:::|||||::::: ||::|||
3 AATCGTGTGCTGGCCGTGCCCTGCCTCG.....GCAGCCGCCCA 46
::: ||||| ::||
36 oValProFaLaCysGlyLySProGLInLeuAsnArGValVaIGlyG 53
::: ||||| ::||
47 CGCGGCCCTCCACCCAGGCGCCCTGCAGCACTGGGCACTGTCGGG 96
::: ||||| ::||
53 IyGIuASerThrAspserGuTrPOTPrPIeValSerILeGLInLyS 69
|||||: ||::|||
97 GTCAAGAGCCCCCAAGCAAGTAGGCTGCAGGTAAAGCTTGAGAGTC 14
::||| |::|||
70 ASngly.....ThHisHCysAlaGlySerLeuLeuThSerAr 83
::||| |::|||

```

147 CAGGCCCATCTGATGCACTTTCGCGGGGCTCCCTCATCCACCCCA 196  
83 gTTPVallIEthrAlaAlaHisCys.....PheLysAsp.... 94  
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
197 GTGGGTGCTGACCGCAGCCGACCTGCTGGACCGGACGTCAAGATCTGG 246  
95 .....AsnLeuAsnLysProTyrLeuPheSerValLeuLeu 106  
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
247 CCGCCCTCAGGGTGCACACTGCGGAGCAGACCTCTAC..... 284  
107 GLyAlaTPrGlnLeuGlnLysnProGlySerArgSerGlnLysValGlyA 123  
285 .....TACCAGGACCAGCTGCTGCCGT 307  
123 lAlaTPrValGlnProHisPro...ValTyrSerTrpLysGlnLysAla 139  
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
308 CAGCAGATCATGCTGCACCCACAGCTTACACCGCCCAAGATCGA.... 353  
139 ySaLaasPllaAlaLeuValArgLeuGlnArgSerIleGlnPheSerGlu 155  
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
54 ..CGGACATCGCCCTGCTGAGCTGAGAGCGCGGTGAACGTCTCCAGC 401  
156 ArgValLeuProIleCysLeuProAspAlaSerIleHisLeuProAs 172  
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
402 CAGCTCCACAGGTCACCTGCCCCCTGCTCAGAGACCTTCCCCCGGG 451  
172 nThrHisCysTrpIleSerGlyTrpGlySerIleGlnAspGlyValPro 189  
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
452 GATCCCGCTGGGTGCTGCTGAGGCGATGTGACAAATGATGAGCGCC 501  
189 euProHisProGlnThrLeuGlnLysLeuLysValProIleIleAspSer 205  
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
502 TCCACCGCATTTCTCTGAAGAGGTGAAGTCCCATATATGGAAGAAC 551  
206 GluValCysSerHisLeuTyrTrpArgGlyAlaGlnGlyPro.... 220  
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
552 CACATTTGTGTGCAAAATACACCTTGGCGCTACACGGGAGACGAGT 601  
221 .....IleThrGlnAspMetLeuCysAlaGlyTyrLeuGlnGlyGlu 235  
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
602 CCGCATGCTCGTGCAGACATGCTGTGCCGG.....AACACCCGGA 645  
235 rGaaspAlaCysLeuGlnLysPserGlyLysProLeuMetCysGlnValAsp 251  
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
646 GGGACTCATGCTCAGGCGGACTCCGAGGCGCCCTGCTGCAAGTGAT 695  
252 GLyAlaTPrLeuLeuAlaGlyIleIleSerTrpGlnGlnCysAlaGln 268  
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
696 GGCACCTGCTGCAGCGGCGGTGCTGAGCGGCGAGGCGGTGTGCCCA 745  
268 uArgAsnArgProGlnValTyrIleSerLeuSerAlaHisArgSerTrp 285  
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
746 GCCCAACCGGCTGCATACCCGCTGCACCTACTGACTGGA 795  
285 aGlnLysIleVal 289  
:|||||  
796 TCCACCACTATGTC 809



XX (GETH ) GENENTECH INC.  
PA Botstein D, Goddard A, Gurney AL, Hillan KJ, Lawrence DA;  
PI Roy MA, Wood WI;  
XX WPI: 1999-430385/36.  
DR N-PSDB: AAX87259.  
XX  
PM Antibody against proteins expressed in neoplastic cells, useful for  
PT tumor diagnosis and treatment  
XX  
PS Example 1; Fig 12; 162pp; English.  
XX  
CC This sequence represents human PRO343 (UNQ302), a protein encoded  
CC by the novel cDNA clone DNA43318 (see AAX87259). Amplification of  
CC DNA43318 was observed in primary lung tumours and in primary colon  
CC tumours, suggesting a significant role in tumour formation and  
CC growth. Antagonists (e.g. antibodies) directed to PRO343 may have  
CC use in cancer therapy. The invention identifies 14 genes (see  
CC AAX87254-67) that are amplified in the genome of tumour cells. Such  
CC amplification is expected to be associated with overexpression of  
CC the gene product and to contribute to tumorigenesis. The encoded  
CC proteins (see AAY06477-90) may be useful targets for the diagnosis  
CC and/or treatment (including prevention) of certain cancers, and may  
CC act as predictors of the prognosis of tumour treatment. Antibodies  
CC that bind the proteins are claimed and used in claimed cancer  
CC diagnostic kits.  
XX  
SQ Sequence 317 AA:  
  
Query Match 100.0%; Score 1720; DB 20; Length 317;  
Best local Similarity 100.0%; Pred. No. 3.2e-147;  
Matches 317; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 MVVSGAPALGGGCGTFTSLTLLASTALINARIPVPACGKPOOLNVGEGDSTSE 60  
DB 1 mvvsgappalgggclgftfslillastallnaaripvpacgkpgqlnvvggedstse 60  
QY 61 WPMIVSIQNGTHHCAGSLTSTRWYITAAHCKDNLNKPYLEFVLLGAMQLGNPSRSQK 120  
DB 61 wpmivsiqngthhcagsllstrwitaahckdnlkpylfsvllgawqlgnpsrsqk 120  
QY 121 VGVAVWEHPHYVSWKEGACADIALVRLERSIOFSERVLPICPDASHLPNTHWISGW 180  
DB 121 vgvawephyvswkegacadalvrlersiofservlpicpdashlpnthcwisgw 180  
QY 181 GSIOGCVPLPHQTLQKLVPIIDSEVCSHLYWRGAGGPTTDMTCAGYLEGERDACIG 240  
DB 181 gsioqcvplphqtlqklvpiidsevcshlywrgaggppttdmtcagylegerdacig 240  
QY 241 DSGGFLMCVOVGAWLTLAIIISMGECACARNRPYVYISIAHSRWEKIVQGVQLRGAQG 300  
DB 241 dsggflmcvgawllagllismgecacrnrpyvylisahsrwekivgvlrgragq 300  
QY 301 GGALRAAPSQSGAARS 317  
DB 301 ggalrapsgsgsaars 317  
  
RESULT 2  
AA13391  
ID AAY13391 standard; Protein: 317 AA.  
XX  
XX AAY13391;  
AC  
XX  
DT 25-JUN-1999 (first entry)  
XX  
DE Amino acid sequence of protein PRO343.  
XX  
DE Secreted protein: transmembrane protein; human; enterocolitis;  
KW Zollinger-Ellison syndrome; gastrointestinal ulceration;  
KW

KW congenital microvillus atrophy; skin disease; cell growth;  
KW abnormal keratinocyte differentiation; psoriasis; epithelial cancer;  
KW Parkinson's disease; Alzheimer's disease; ALS; neuropathy;  
KW fibromodulin; dermal scarring; Usher Syndrome; Atrophia areata;  
KW anti-thrombotic; wound healing; tissue repair.  
XX  
XX Homo sapiens.  
XX  
XX  
XX PN W09914328-A2.  
XX PD 25-MAR-1999.  
XX  
XX  
XX PF 16-SEP-1998; 98WO-US19330.  
XX  
XX PR 25-NOV-1997; 97US-0066840.  
XX PR 17-SEP-1997; 97US-0059113.  
XX PR 17-SEP-1997; 97US-0059115.  
XX PR 17-SEP-1997; 97US-0059117.  
XX PR 17-SEP-1997; 97US-0059119.  
XX PR 17-SEP-1997; 97US-0059121.  
XX PR 17-SEP-1997; 97US-0059122.  
XX PR 18-SEP-1997; 97US-0059184.  
XX PR 18-SEP-1997; 97US-0059263.  
XX PR 15-SEP-1997; 97US-0059266.  
XX PR 15-OCT-1997; 97US-0062125.  
XX PR 17-OCT-1997; 97US-0062285.  
XX PR 21-OCT-1997; 97US-0062287.  
XX PR 24-OCT-1997; 97US-0063486.  
XX PR 24-OCT-1997; 97US-0062814.  
XX PR 24-OCT-1997; 97US-0062816.  
XX PR 24-OCT-1997; 97US-0063045.  
XX PR 24-OCT-1997; 97US-0063120.  
XX PR 24-OCT-1997; 97US-0063121.  
XX PR 24-OCT-1997; 97US-0063127.  
XX PR 24-OCT-1997; 97US-0063128.  
XX PR 27-OCT-1997; 97US-0063329.  
XX PR 27-OCT-1997; 97US-0063327.  
XX PR 28-OCT-1997; 97US-0063341.  
XX PR 28-OCT-1997; 97US-0063542.  
XX PR 28-OCT-1997; 97US-0063544.  
XX PR 28-OCT-1997; 97US-0063549.  
XX PR 28-OCT-1997; 97US-0063550.  
XX PR 28-OCT-1997; 97US-0063564.  
XX PR 29-OCT-1997; 97US-0063732.  
XX PR 29-OCT-1997; 97US-0063730.  
XX PR 29-OCT-1997; 97US-0063738.  
XX PR 29-OCT-1997; 97US-0063734.  
XX PR 29-OCT-1997; 97US-0064215.  
XX PR 29-OCT-1997; 97US-0063735.  
XX PR 31-OCT-1997; 97US-0063870.  
XX PR 31-OCT-1997; 97US-0064103.  
XX PR 03-NOV-1997; 97US-0064248.  
XX PR 07-NOV-1997; 97US-0064809.  
XX PR 12-NOV-1997; 97US-0065186.  
XX PR 17-NOV-1997; 97US-0065846.  
XX PR 18-NOV-1997; 97US-0065693.  
XX PR 21-NOV-1997; 97US-0066120.  
XX PR 21-NOV-1997; 97US-0066364.  
XX PR 24-NOV-1997; 97US-0066772.  
XX PR 24-NOV-1997; 97US-0066466.  
XX PR 24-NOV-1997; 97US-0066770.  
XX PR 24-NOV-1997; 97US-0066511.  
XX PR 24-NOV-1997; 97US-0066453.  
  
(GETH ) GENENTECH INC.  
XX  
XX  
XX PA Chen J, Goddard A, Gurney AL, Pennica D, Wood WI, Yuan J;  
XX PI WPI: 1999-229533/19.  
XX DR N-PSDB: AAX52262.  
XX  
XX New isolated human genes and polypeptides used in, e.g. treatment of

PT	gastrointestinal ulceration
XX	
PS	Claim 12; Fig 98; 320pp; English

**SQ** Sequence 317 AA;

Query Match	100.0%	Score 1720;	DB 20;	Length 317;
Best Local Similarity	100.0%	Pred. No. 3	2e-147;	
Matches 317; Conservative	0;	Mismatches	0;	Indels 0; Gaps 0;

QY 1 MVSAPPALGGGCLTFTSLILASTAILNARIIPVPACGKPGQLNRVVGEDSDISE 60  
|||  
Db 1 mvsgappalggclygtftslilastailnaaripvpacgkpgqlnrvvgedsdise 60

QY 61 WPMIVSIQKNGTGHHCASLLISRWVIVAAHCFKUNLNKPLESVLLGAWQJGAFPGUSISQN 120

Db 61 WPMIVSIQKNGTGHHCASLLISRWIVTAHCFKdnlkpylfsvllgawqjgnpgsrssqk 120

Db 121 vgvawerphpyvswkegacadaiaivrlersiqfservlpriclpdasihlprntncwlsqw 180

181 gsiqgvp1pnpqtlqrlkvpildseweshlywrgagqprltemlcaagylegerdacly 240

Db 241 dsgrpmcqvawllaglliswgegaernrpyylslahrswekivgavqlrragq 300

Db 301 ggalrapsqsgsaaars 317

RESULT	3
AB11700	
ID	AB11700 standard; Protein; 317 AA.

AC	AAB11700;
XX	
DT	23-OCT-2000 (first entry)
VV	

DE Human serine protease BSSP4 (MBSSP4) SEQ ID NO.2.  
XX  
XX BSSP4; serine protease; human; hBSSP4; brain;  
KW diagnostic marker; antibody; transgenic animal; Alzheimer's disease;

AM	XX	OS
cedentia, allops?	cancel	intimidation/
		Freedom/
		cancel/
		cancel/

XX WO200031277-A1  
PN

PD	02-JUN-2000.
XX	
PF	19-NOV-1999; 99WO-JP06472.
XX	

PR 20-NOV-1998; 98JF-034/813.  
XX  
PA (FUSO ) FUSO PHARM IND LTD.  
XX

PI Denial n, Oral n, Nonintentional  
XX  
DR WPI; 2000-400084/34.  
XX

PT	diagnosis of diseases in which
XX	Claim 1; Page 66-67; 111pp;
PS	

CC The invention relates to novel  
CC (AAB11700-B11709), and to novel  
CC AAA61799). The invention also

CC expression level of BSSP4 cat  
CC The invention additionally  
CC of production of such antib

CC markers for certain medical  
CC initially isolated in a huma  
CC primers (AAA61714-A61715) b  
CC The pSP4 serine

CC useful in detecting homologous  
CC biological samples (e.g., blood  
CC as diagnostic markers for disease  
CC expression levels. Such dis-

CC (diapry), cancer of the bladder  
CC Sequences AAA61695-A61703 a  
CC BSSP4 variants (hBSSP4), an  
CC murine BSSP4 (mBSSP4). Sequ

CC Variance (mbssp4) .  
CC (mbssp4) .  
XX  
SQ Sequence 317 AA;

Query Match	100.0%
Best Local Similarity	100.0%

QY 1 M V S G A P A L G G C L G T F T S L L  
| | | | | | | | | | | | | | | |  
1 1 1 1 1 1 1 1 1 1 1 1 1 1 1

61 WPWIVSIQKNGTHHCAGSLTTS  
|||

QY	121	VCVAMVEPVPVYSWKEGACADII
Rb	121	VCVAMVEPVPVYSWKEGACADII

QY	181	GSIQDGVPLRHPQTLLQKLKVPIT
Dh	181	asidavpplphpatlaklkvpri

OY	241	DSCGPLMCQVDGAWLLAGIISW 
Dd	241	dsqgplmcqvdaqwl laqilisw

QY	301	GGALRAPSQSGGAARS	317
Db	301	ggalrapsgsgsaaars	317

```

RESULT 4
AAV93689
ID AAV93689 standard; Protein; 317 AA.
XX
AC AAV93689;
XX
DT 03-OCT-2000 (first entry)
XX
DE Amino acid sequence of novel polypeptide PRO343.
XX
KW PRO201; PRO292; PRO327; PRO1265; PRO344; PRO343; PRO347; PRO357;
KW PRO715; PRO1017; PRO1112; PRO509; PRO853; PRO882; tumour cell;
KW tumorigenesis; cancer; neoplastic cell growth; cell proliferation.
XX
OS Homo sapiens.
XX
FT Key
FT Peptide
FT Location/Qualifiers
FT 1..32
FT /note= "signal sequence"
FT Region
FT 4..15
FT /note= "prokaryotic membrane lipoprotein attachment site"
FT Modified-site
FT 5..11
FT /note= "N-myristoylation site"
FT Modified-site
FT 12..18
FT /note= "N-myristoylation site"
FT Modified-site
FT 13..19
FT /note= "N-myristoylation site"
FT Modified-site
FT 16..22
FT /note= "N-myristoylation site"
FT Modified-site
FT 52..58
FT /note= "N-myristoylation site"
FT Modified-site
FT 70..74
FT /note= "N-glycosylation site"
FT Modified-site
FT 71..77
FT /note= "N-myristoylation site"
FT Modified-site
FT 77..83
FT /note= "N-myristoylation site"
FT Active-site
FT 86..92
FT /note= "active site for serine protease"
FT Modified-site
FT 112..118
FT /note= "N-myristoylation site"
FT Region
FT 178..182
FT /note= "glycosylation attachment site"
FT Modified-site
FT 273..279
FT /note= "N-myristoylation site"
FT Modified-site
FT 310..316
FT /note= "N-myristoylation site"
MO200037640-A2.
XX
PD 29-JUN-2000.
XX
PF 16-DEC-1999; 99WO-US30095.
XX
PR 22-DEC-1998; 98US-0113296.
PR 08-MAR-1999; 99WO-US05028.
PR 02-JUN-1999; 99WO-US12252.
PR 01-SEP-1999; 99WO-US20111.
PR 15-SEP-1999; 99WO-US21090.
PR 30-NOV-1999; 99WO-US28313.
PR 01-DEC-1999; 99WO-US28409.
PR 02-DEC-1999; 99WO-US28301.
PR 02-DEC-1999; 99WO-US28565.
XX
PA (GENET) GENENTECH INC.
XX
PI Botstein D, Goddard A, Gurney AL, Hillan K, Lawrence DA, Roy MA;
PI Wood WT;
XX
DR WPI; 2000-452188/39.
DR N-PSDB; AAA46914.

```

```

XX
PT New anti-polypeptide antibody useful in the treatment and diagnosis of
PT neoplastic cell growth and proliferation -
XX
XX Claim 61; Fig 12; 220pp; English.
XX
PS
XX
CC The present sequence represents a novel human polypeptide. The
CC specification describes novel polypeptides designated PRO201, PRO292,
CC PRO327, PRO1265, PRO344, PRO343, PRO347, PRO357, PRO715, PRO1017,
CC PRO1112, PRO509, PRO853 and PRO882. These genes are amplified in
CC the genome of tumour cells. The polypeptides are believed to contribute
CC to tumorigenesis. The polypeptides are useful target for the
CC identification of certain cancers, and may act as predictors of the
CC prognosis of tumour treatment. Antibodies against these polypeptides
CC are useful in the treatment and diagnosis of neoplastic cell growth
CC and proliferation in mammals.
XX
SQ Sequence 317 AA:
XX
Query Match 100.0%; Score 1720; DB 21; Length 317;
Best Local Similarity 100.0%; Pred. No. 3.2e-147;
Matches 317; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MVSAGPALGGGCLGTFSTLLSLASTALNARIPVPACGKPOLNRVVGEDSTDSE 60
DB 1 mvvsagpalgggclgtftslllslastalnaaripvpacgkpgqlnrvggedstse 60
QY 61 WMYVSIQNGNTHRCAGSLLTSRWYITAAHCKKDKMLKPYLFSLVGLGAMOLGNPGRSQK 120
DB 61 wmyvsiqngnthrcagslltsrwyitaa hckkdkmlkpylfsvlvgwqlgnpgrsqk 120
QY 121 VGVAVWEPPHYVSWKEGACADIALVRLERSIOFSERVPICLPDASIHLPNTHCWISGW 180
DB 121 vgvawvepphyvswkegacadiavrlersiofservpiclpdasihlpnthcwisgw 180
QY 181 GSIDGVPLPHPQTLQKLVPIIDSEVCSHLVWRGAGGPTEDMLCAGYLEGERDACLG 240
DB 181 gsidgvplphpqtlqklvpiidsevcshlvwrgaggptedmlcagylegerdaclg 240
QY 241 DSGGPLMCQVOCAMWLAGIISWEGECARNPQYIISARSWEKIVQGVOLRGAROG 300
DB 241 dsggplmcqvocawlwagiiswgegcaernpqyisarswekivqgvqlrgrag 300
QY 301 GGALRAPSGSGGAARS 317
DB 301 ggalrapsgsgsaars 317
RESULT 5
AAV72890
ID AAV72890 standard; Protein; 317 AA.
XX
AC AAV72890;
XX
DT 31-MAY-2001 (first entry)
XX
DE Human serine protease, protease C-E.
XX
DE Human serine protease, protease C-E.
XX
KW Human; serine protease; protease C-E; therapy; desquamation; skin care;
KW laundry detergent; shampoo; cleaning agent; hair care; skin flaking;
KW neurodegenerative disorder; dermatological; immunogenic; proteolytic.
KW Chromosome 16p13.3.
XX
OS Homo sapiens.
XX
OS
XX
PN WO200116288-A2.
XX
PD 08-MAR-2001.
XX
PF 14-AUG-2000; 2000WO-US22117.
XX
PR 31-AUG-1999; 99US-0386629.

```



XX (ORTH ) ORTHO-MCNETL PHARM INC.  
 XX PA  
 XX PI Darrow A, Qi J, Andrade-Gordon P;  
 XX DR WPI; 2001-226681/23.  
 XX DR N-PSDB; AAD0290.  
 XX PT Novel serine protease termed protease C-E, useful for treating and  
 XX PT preventing skin flaking or imbalance of desquamation -  
 XX PS Claim 11; Fig 1; 78pp; English.  
 CC The present sequence is a human serine protease, protease C-E which  
 CC is a member of the SI serine protease family. Protease C-E gene is  
 CC located on chromosome 16p13.3 and is expressed in pancreas, placenta,  
 CC prostate, small intestine, stomach, spleen, fibroblasts, epidermis,  
 CC cerebellum, cerebral cortex, pituitary and hippocampus. Protease C-E is  
 CC useful for treating an imbalance of desquamation, by topical application.  
 CC A non-pharmaceutical composition comprising the protein may be formulated  
 CC as a laundry detergent, shampoo, hard surface cleaning composition, dish  
 CC care cleaning composition, skin care composition and hair care  
 CC composition. Protease C-E is useful for treating and preventing skin  
 CC flaking, neurodegenerative disorders and dermatological pathologies. It  
 CC is less immunogenic to sensitive individuals and it provides efficient  
 CC proteolytic activity in a non-natural environment.  
 XX SQ Sequence 317 AA;

Query Match 100.0%; Score 1720; DB 22; Length 317;  
 Best Local Similarity 100.0%; Pred. No. 3.2e-147;  
 Matches 317; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MYVSGAPPALGGCGCTFTSLILASTALINAAIRPVPPACGKPOQLNRVVGEDSTDSE 60  
 DB 1 MYVSGAPPALGGCGCTFTSLILASTALINAAIRPVPPACGKPOQLNRVVGEDSTDSE 60  
 OY 61 WPWIVSIQKNGTHHCAGSLTSRWVITTAHCFKDNLNKPYLFSVILGAMOLGNPGRSROK 120  
 DB 61 WPWIVSIQKNGTHHCAGSLTSRWVITTAHCFKDNLNKPYLFSVILGAMOLGNPGRSROK 120  
 OY 121 VGVAMVEHPHYVSWKEGACADIALVRLERSIQSESRVLPICLPDASIHLPNTHCWTISGW 180  
 DB 121 VGVAMVEHPHYVSWKEGACADIALVRLERSIQSESRVLPICLPDASIHLPNTHCWTISGW 180  
 OY 121 VGVAMVEHPHYVSWKEGACADIALVRLERSIQSESRVLPICLPDASIHLPNTHCWTISGW 180  
 DB 121 VGVAMVEHPHYVSWKEGACADIALVRLERSIQSESRVLPICLPDASIHLPNTHCWTISGW 180  
 OY 181 GSIDQGVPLPHPQTLQKLIKVPFIIDSEVCSHLYWKGAGCGPTREEMLCAGVLEGGRDCLG 240  
 DB 181 GSIDQGVPLPHPQTLQKLIKVPFIIDSEVCSHLYWKGAGCGPTREEMLCAGVLEGGRDCLG 240  
 OY 241 DSGGPLMCOVDGAWLAGIISWEGGCAERNRPVYISLSAHSRWKVKIVGVOLRGRAGC 300  
 DB 241 DSGGPLMCOVDGAWLAGIISWEGGCAERNRPVYISLSAHSRWKVKIVGVOLRGRAGC 300  
 OY 301 GGALRAPSGGGAARS 317  
 DB 301 GGALRAPSGGGAARS 317  
 OY 301 GGALRAPSGGGAARS 317  
 DB 301 GGALRAPSGGGAARS 317

RESULT 6  
 AAB80259  
 ID AAB80259 standard; Protein; 317 AA.  
 XX AC AAB80259;

XX 24-APR-2001 (first entry)  
 XX DE Human PRO343 protein.  
 XX Human PRO343 protein.  
 KW Human: PRO: dermatological; antipsoriatic; cytostatic; antiinflammatory;  
 KW antiParkinsonian neurotropic; neuroprotective; vulnereary; cardiatic;  
 KW antiangiogenic; vasotrophic; antiasthmatic; antihypertensive; cancer;  
 KW antiarthritic; antifertility; antidiabetic; antiviral; diabetes;

KW ophthalmological; gene therapy; skin disease; gastrointestinal disorder;  
 KW ischaemia; inflammation.  
 XX OS Homo sapiens.  
 XX PN WO200104311-A1.  
 XX PD 18-JAN-2001.  
 XX PF 22-FEB-2000; 2000WO-US04414.  
 XX PR 07-JUL-1999; 99US-0143048.  
 XX PR 26-JUL-1999; 99US-0143698.  
 XX PR 28-JUL-1999; 99US-0146222.  
 XX PR 08-SEP-1999; 99WO-US20594.  
 XX PR 13-SEP-1999; 99WO-US20944.  
 XX PR 15-SEP-1999; 99WO-US21090.  
 XX PR 15-SEP-1999; 99WO-US21547.  
 XX PR 05-OCT-1999; 99WO-US23089.  
 XX PR 29-NOV-1999; 99WO-US28214.  
 XX PR 30-NOV-1999; 99WO-US28313.  
 XX PR 16-DEC-1999; 99WO-US30095.  
 XX PR 20-DEC-1999; 99WO-US30911.  
 XX PR 20-DEC-1999; 99WO-US30999.  
 XX PR 05-JAN-2000; 99WO-US00219.  
 XX (GENTH ) GENTECH INC.

PI Ashkenazi AJ, Botstein D, Desnoyers L, Eaton DL, Ferrara N;  
 PI Filvaroff E, Fong S, Go W, Gerber H, Gerritsen ME, Goddard A;  
 PI Godowski PJ, Grimaldi CJ, Gurney AL, Hillian KJ, Kljavin JT;  
 PI Mather JP, Pan J, Paoni NF, Roy MA, Stewart TA, Tumas D;  
 PI Williams PM, Wood WI;  
 DR WPI; 2001-081051/09.  
 DR N-PSDB; AAF72420.  
 XX

PT Sixty one nucleic acids encoding PRO polypeptides which are useful in  
 PT the treatment of skin diseases (e.g. psoriasis), cancers (e.g. lung  
 PT squamous cell carcinoma) and neurodegenerative diseases (e.g.  
 PT Alzheimer's disease) -  
 XX

Claim 1; Fig 98; 393pp; English.

CC The present sequence is one of sixty one novel secreted and  
 CC transmembrane PRO polypeptides. The PRO polypeptides are  
 CC useful for treating skin diseases (e.g. psoriasis), cancers (e.g. lung  
 CC squamous cell carcinoma), gastrointestinal disorders (e.g.  
 CC enterocolitis), neurodegenerative diseases (e.g. Alzheimer's disease,  
 CC Parkinson's disease), wound repair, cardiovascular disorders (e.g.  
 CC endometrial bleeding, angiodysplasia, ischaemias such as coronary  
 CC ischaemia, atherosclerosis), inflammatory disorders (e.g. asthma,  
 CC rheumatoid arthritis, multiple sclerosis), infertility, AIDS and  
 CC diabetes and retinal disorders such as retinitis pigmentosa.  
 CC The PRO nucleic acids have applications in molecular biology, including  
 CC use as hybridization probes, and in chromosome and gene mapping.  
 XX

SQ Sequence 317 AA;

Query Match 100.0%; Score 1720; DB 22; Length 317;  
 Best Local Similarity 100.0%; Pred. No. 3.2e-147;  
 Matches 317; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MYVSGAPPALGGCGCTFTSLILASTALINAAIRPVPPACGKPOQLNRVVGEDSTDSE 60  
 DB 1 MYVSGAPPALGGCGCTFTSLILASTALINAAIRPVPPACGKPOQLNRVVGEDSTDSE 60  
 OY 61 WPWIVSIQKNGTHHCAGSLTSRWVITTAHCFKDNLNKPYLFSVILGAMOLGNPGRSROK 120  
 DB 61 WPWIVSIQKNGTHHCAGSLTSRWVITTAHCFKDNLNKPYLFSVILGAMOLGNPGRSROK 120  
 OY 121 VGVAMVEHPHYVSWKEGACADIALVRLERSIQSESRVLPICLPDASIHLPNTHCWTISGW 180  
 DB 121 VGVAMVEHPHYVSWKEGACADIALVRLERSIQSESRVLPICLPDASIHLPNTHCWTISGW 180

Db 121 ygvawephpyswkegacadiatvrlersigfseervlplclpasihlpnthcwisgw 180  
QY 181 GSIDGVLPHPQTLQKAVIITSEVCSHLYMRAGGPTTEMLCAGYEGEGRDCLG 240  
Db 181 gsidgvlpHPqtlqKAvIITSEVCSHLYMRAGGPTTEMLCAGYEGEGRDCLG 240  
QY 241 DSGGFLMCQVGGAMLGIIISWEGSCAERNRPVYIISLAHRSWEKIVQGVOLRGRAQG 300  
Db 241 dsggflmcqvGGaMLlGIIISWEGSCAERNRPVYIISLAHRSWEKIVGvQLrGrAg 300  
QY 301 GGALRAPSGSGGAARS 317  
Db 301 ggalrapsgsgsGaars 317  
RESULT 7  
AAU23215  
AAU23215 standard; Protein: 325 AA.  
AAU23215;  
17-DEC-2001 (first entry)  
XX Novel human enzyme polypeptide #301.  
XX  
KW Human; oxidoreductase enzyme; transferase; hydrolase; lyase; isomerase;  
KW ligase; hyperproliferative disorder; immunodeficiency disorder;  
KW autoimmune disorder; neurological disorder; metabolic disorder;  
KW inflammatory disorder; cardiovascular disorder; reproductive disorder;  
KW blood-related disorder; infectious disorder; cytostatic; anti arthritic;  
KW nephrotropic; anticoagulant.  
XX  
OS Homo sapiens.  
XX  
PN WO200155301-A2.  
XX  
PD 02-AUG-2001.  
XX  
PF 17-JAN-2001; 2001WO-US01239.  
XX  
XX 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 27-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226661.

PR 22-AUG-2000; 2000US-0226668.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229513.  
PR 05-SEP-2000; 2000US-0229519.  
PR 05-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 21-SEP-2000; 2000US-0234977.  
PR 25-SEP-2000; 2000US-0234998.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239355.  
PR 13-OCT-2000; 2000US-0239357.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.

PI	Rosen CA,	Barash SC,	Ruben SM,
XX			
DR	WPI; 2001-465566/50.		
DR	N-PSDB; AAS41085.		

Claim 11; SEQ ID No 1211; 1180bp; English

Sequence 325 AA;

ry Match	100.0%;	Score 1720;	DB 22;	Length 325;
t Local Similarity	100.0%;	Pred. No. 3.3e-147;		
chases 317; Conservative	0;	Mismatches	0;	Indels 0; Gaps 0

Db 309 gga|rapsggsgaaars 325

ID	AAU17037	standard; 325 AA
YX		

AC AAU17037;

DT 07-NOV-2001 (first entry)

Human novel secreted protein, SEQ ID 278

KM	hypertension
KM	hypotension
KM	immunosuppressive
KM	antiarthritic
KM	antirheumatic
KM	antitumor
KM	cytostatic
KM	cardiant
KM	vasotropic
KM	cerebroprotective
KM	neuroprotective
KM	antibacterial
KM	virucide
KM	fungicide
KM	ophthalmological
KM	vulnary
KM	secreted protein
KM	rheumatoid arthritis
KM	hyperproliferative disorder
KM	cardiovascular disorder
KM	cardiac arrest
KM	cerebrovascular disorder
KM	cerebral ischemia
KM	angiogenesis
KM	nervous system disorder
KM	Alzheimer's disease
KM	infection
KM	ocular disorder
KM	corneal infection
KM	wound healing
KM	epithelial cell proliferation
KM	skin ageing
KM	food additive
KM	preservative
KM	antiproliferative

Homo sapiens

PN WO200155441-A2

PD 02-AUG-2001.

PF 17-JAN-2001; 2001WO-US01320.

PR	31-JAN-2000;	2000US-0178065.
PR	04-FEB-2000;	2000US-0180628.
PR	24-FEB-2000;	2000US-0184664.
PR	02-MAR-2000;	2000US-0186350.
PR	16-MAR-2000;	2000US-0189874.
PR	17-MAR-2000;	2000US-0190076.
PR	18-APR-2000;	2000US-0198123.
PR	19-MAY-2000;	2000US-0205515.
PR	07-JUN-2000;	2000US-0209467.
PR	28-JUN-2000;	2000US-0214886.
PR	30-JUN-2000;	2000US-0215135.
PR	07-JUL-2000;	2000US-0216647.
PR	07-JUL-2000;	2000US-0216680.
PR	11-JUL-2000;	2000US-0217487.
PR	11-JUL-2000;	2000US-0217486.
PR	14-JUL-2000;	2000US-0218290.
PR	26-JUL-2000;	2000US-0220963.
PR	26-JUL-2000;	2000US-0220964.
PR	14-AUG-2000;	2000US-0224518.
PR	14-AUG-2000;	2000US-0224519.



disorders listed in the specification. The polypeptides can also be used to aid wound healing and epithelial cell proliferation, to prevent skin aging due to sunburn, to maintain organs before transplantation, for supporting cell culture of primary tissues, to regenerate tissues and in chemotaxis. The polypeptides can also be used as a food additive or preservative to increase or decrease storage capabilities, fat content, lipid, protein, carbohydrate, vitamins, minerals, cofactors and other nutritional components. The present

Query Match 100.0%; Score 1720; DB 22; Length 325;  
Best Local Similarity 100.0%; Pred No.3.e-147;  
Matches 317; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 MVVSGAPPALGGGCLGTFSTLLASTATLAAARIPVPAKGRQOLNRYVGSGEDSTDE 60  
|||||  
9 mvvsgappalgggclgftstllastatlnaariipvpacgkpgqlnrvvggedstde 68  
61 WPIVTSIQKNGTHHAGSLTSTWVITAAHCFDNLNKPFLFVLLGAWQLGMPGSRQK 120  
|||||  
69 wpiwtsiqkngthhacgsltstwvitaahcfkdnlnkpyllfgawqlgmpgsrsgk 128  
121 VGVAVWEPHPVYSWKEGACADIALVRLERSIQFSERVLPICLPDASIHLPNTHCWISGW 180  
|||||  
129 vgvawephpvyswkegacadiavrlersiqfservlpiclpdasihlpnchwisgw 188  
181 GSIDGVPLPHTQTLQKIKVPIIDSEVCSHLVWRAGCGPITEDMLCAGLGEGRDACLG 240  
|||||  
189 gsidgvplphtqtlqkikvpiidsevcshlywrgagqgpltdmlcagylegerdclg 248  
241 DSGGFLMCQVNGAWLLAGITISWEGCAERNRPYVYISLSHRSWVEKIVQGYOLRGRAQG 300  
|||||  
249 dsggflmcvngawllagitlswegcaernpvyylsishrswekivqgyolrgragq 308  
301 GGALRAPSOGSGAARS 317  
|||||  
309 ggalrapsogsgaars 325

RESULT 9  
AAU16966  
ID AUI16966 standard; Protein; 351 AA.  
XX  
AC AUI16966;  
XX  
DT 07-NOV-2001 (first entry)  
XX  
DE Human novel secreted protein, SEQ ID 207.  
Human; immunosuppressive; antiarthritic; antirheumatic;  
cytostatic; cardiant; vasotropic; cerebroprotective; nootropic;  
neuroprotective; antibacterial; virocidic; fungicide; optalmatological;  
vulnerary; secreted protein; rheumatoid arthritis;  
hyperproliferative disorder; cardiovascular disorder; angiodenesis;  
cerebrovascular disorder; cerebral ischaemia; angiogenesis;  
nervous system disorder; Alzheimer's disease; infection; ocular disorder;  
corneal infection; wound healing; epithelial cell proliferation;  
skin ageing; food additive; preservative; antiproliferative.

OS Homo sapiens.  
XX  
XX  
PN WO200155441-A2.  
XX  
PD 02-AUG-2001.  
XX  
PF 17-JAN-2001; 2001WO-US01320.  
XX  
XX 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0188874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.

PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218299.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226688.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227099.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0228287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.

Query	Match	100.0%	Score 1720;	DB 22;	Length 351;
Best Local Similarity	100.0%	Pred. No. 3.6e-147;			
Matches 317;	Conservative	0;	Mismatches 0;	Indels 0;	Gaps 0
QY	1	MVSGAPALPAGGCTCTFTSLLLASTALTNARIPVPACGKPOOLNRVVGEDSTDSE	60		
Db	35	mvsgappalpaggcgtctftslilastalnariypvacgkpgqlnrvggedstdse	94		
QY	61	WPWIVSIQKNGTHHCAGSLTISRMYITAAHCKRDLNKPYLESVLGAQMDGNBSRSQK	120		
Db	95	wpwivsqtqkgtthhcagslltsrmyitaaahckrdlnkpylfsvllgawqgnpgrsqtq	154		
QY	121	VGVAAWVEPHVYVSWKRGACADIALVRLERSTQFSRVLPICLPASHTLPNTHCWISGW	180		
Db	155	vgvawvephvpyvswkrgacadiavrlersqtfsrvlpiclpashtlpnthcwisgw	214		
QY	181	GSIQDGVPLPHQDTLOKTLKVPRIIDSEVCSHLYWRGAGOGPFTTEMLCAGYLEGERDACLQ	240		
Db	215	gsiqdgvplphqdtlqklkvpriidsevcshlywrgagogpfttedmlcagylegerdacig	274		
QY	241	DGSGPLMCOVDGAWMLAGIISWGECARPNRPVYISLSAHSVKEITVQGVOLRGAAQS	300		
Db	275	dsgsplmcovdgawmlagiiswgecaerpnrvyislsahsvkeitvqgvqltrragq	334		
QY	301	GGALRAPSGSGGAARSS 317			
Db	335	ggalrapsgsgsaaars 351			
RESULT 10					
ID	AAU23751				
	AAU23751 standard; Protein: 322 AA.				
AC	AAU23751;				
XX	18-DEC-2001 (first entry)				
XX	Novel human enzyme polypeptide #837.				
XX	Human: oxidoreductase enzyme; transferase; hydrolase; lyase; isomerase;				
KW	lignase; hyperproliferative disorder; immunodeficiency disorder;				
KW	autoimmune disorder; neurological disorder; metabolic disorder;				
KW	inflammatory disorder; cardiovascular disorder; reproductive disorder;				
KW	blood-related disorder; infectious disorder; cytostatic; anti arthritic;				
KW	nephrotropic; anticoagulant.				
OS	Homo sapiens.				
XX	WO200155301-A2.				
XX	02-AUG-2001.				

Page 11

XX	17-JAN-2001; 2001WO-US01239.	PR	29-SEP-2000; 2000US-0236369.
XX		PR	29-SEP-2000; 2000US-0236370.
PR	31-JAN-2000; 2000US-0179065.	PR	02-OCT-2000; 2000US-0236802.
PR	04-FEB-2000; 2000US-0180628.	PR	02-OCT-2000; 2000US-0237037.
PR	24-FEB-2000; 2000US-0184664.	PR	02-OCT-2000; 2000US-0237038.
PR	02-MAR-2000; 2000US-0186350.	PR	02-OCT-2000; 2000US-0237039.
PR	16-MAR-2000; 2000US-0186874.	PR	02-OCT-2000; 2000US-0237040.
PR	17-MAR-2000; 2000US-0190076.	PR	11-OCT-2000; 2000US-0239935.
PR	18-APR-2000; 2000US-0198123.	PR	11-OCT-2000; 2000US-0239937.
PR	19-MAY-2000; 2000US-0205515.	PR	20-OCT-2000; 2000US-0240960.
PR	07-JUN-2000; 2000US-0209457.	PR	20-OCT-2000; 2000US-0241121.
PR	28-JUN-2000; 2000US-0214886.	PR	20-OCT-2000; 2000US-0241785.
PR	30-JUN-2000; 2000US-0215135.	PR	20-OCT-2000; 2000US-0241786.
PR	07-JUL-2000; 2000US-0216647.	PR	20-OCT-2000; 2000US-0241787.
PR	07-JUL-2000; 2000US-0216880.	PR	20-OCT-2000; 2000US-0241808.
PR	11-JUL-2000; 2000US-0217487.	PR	20-OCT-2000; 2000US-0241809.
PR	14-JUL-2000; 2000US-0218290.	PR	01-NOV-2000; 2000US-0241826.
PR	06-JUL-2000; 2000US-0220953.	PR	08-NOV-2000; 2000US-0246474.
PR	14-AUG-2000; 2000US-0220964.	PR	08-NOV-2000; 2000US-0246475.
PR	14-AUG-2000; 2000US-0224518.	PR	08-NOV-2000; 2000US-0246476.
PR	14-AUG-2000; 2000US-0224519.	PR	08-NOV-2000; 2000US-0246477.
PR	14-AUG-2000; 2000US-0225213.	PR	08-NOV-2000; 2000US-0246478.
PR	14-AUG-2000; 2000US-0225214.	PR	08-NOV-2000; 2000US-0246523.
PR	14-AUG-2000; 2000US-0225266.	PR	08-NOV-2000; 2000US-0246524.
PR	14-AUG-2000; 2000US-0225267.	PR	08-NOV-2000; 2000US-0246525.
PR	14-AUG-2000; 2000US-0225268.	PR	08-NOV-2000; 2000US-0246526.
PR	14-AUG-2000; 2000US-0225270.	PR	08-NOV-2000; 2000US-0246527.
PR	14-AUG-2000; 2000US-0225447.	PR	08-NOV-2000; 2000US-0246528.
PR	14-AUG-2000; 2000US-0225757.	PR	08-NOV-2000; 2000US-0246532.
PR	14-AUG-2000; 2000US-0225758.	PR	08-NOV-2000; 2000US-0246609.
PR	18-AUG-2000; 2000US-0225759.	PR	08-NOV-2000; 2000US-0246610.
PR	22-AUG-2000; 2000US-0226279.	PR	08-NOV-2000; 2000US-0246611.
PR	22-AUG-2000; 2000US-0226681.	PR	08-NOV-2000; 2000US-0246613.
PR	22-AUG-2000; 2000US-0226686.	PR	17-NOV-2000; 2000US-0249207.
PR	22-AUG-2000; 2000US-0227182.	PR	17-NOV-2000; 2000US-0249208.
PR	23-AUG-2000; 2000US-0227009.	PR	17-NOV-2000; 2000US-0249209.
PR	30-AUG-2000; 2000US-0228924.	PR	17-NOV-2000; 2000US-0249210.
PR	01-SEP-2000; 2000US-0229827.	PR	17-NOV-2000; 2000US-0249211.
PR	01-SEP-2000; 2000US-0229843.	PR	17-NOV-2000; 2000US-0249212.
PR	01-SEP-2000; 2000US-0229345.	PR	17-NOV-2000; 2000US-0249213.
PR	05-SEP-2000; 2000US-0229509.	PR	17-NOV-2000; 2000US-0249215.
PR	06-SEP-2000; 2000US-0230437.	PR	17-NOV-2000; 2000US-0249217.
PR	06-SEP-2000; 2000US-0230438.	PR	17-NOV-2000; 2000US-0249218.
PR	08-SEP-2000; 2000US-0231242.	PR	17-NOV-2000; 2000US-0249244.
PR	08-SEP-2000; 2000US-0231243.	PR	17-NOV-2000; 2000US-0249245.
PR	08-SEP-2000; 2000US-0231244.	PR	17-NOV-2000; 2000US-0249264.
PR	08-SEP-2000; 2000US-0231413.	PR	17-NOV-2000; 2000US-0249265.
PR	08-SEP-2000; 2000US-0231414.	PR	17-NOV-2000; 2000US-0249295.
PR	08-SEP-2000; 2000US-0232080.	PR	17-NOV-2000; 2000US-0249299.
PR	12-SEP-2000; 2000US-0233081.	PR	17-NOV-2000; 2000US-0249300.
PR	14-SEP-2000; 2000US-0233966.	PR	01-DEC-2000; 2000US-0250160.
PR	14-SEP-2000; 2000US-0233967.	PR	01-DEC-2000; 2000US-0250331.
PR	14-SEP-2000; 2000US-0233968.	PR	05-DEC-2000; 2000US-0251030.
PR	14-SEP-2000; 2000US-0233969.	PR	05-DEC-200

XX Novel polypeptides and polynucleotides useful for diagnosing,  
 PT preventing, treating neural, immune system, muscular, reproductive,  
 PT pulmonary, cardiovascular, renal, proliferative disorders and cancerous  
 PT diseases -

PS Claim 11: SEQ ID No 1747; 1180bp; English.

XX  
 CC The present invention relates to the isolation of novel human enzyme  
 CC polypeptides, and the cDNA (AA040785-AA041684) and genomic sequences  
 CC encoding them. The enzyme polypeptides of the invention may comprise the  
 CC functional classes of oxidoreductases, transferases, hydrolases, lyases,  
 CC isomerases or ligases. The sequences of the invention are useful in the  
 CC diagnosis, treatment, prevention and/or prognosis of a wide range of  
 CC disorders, including hyperproliferative disorders (e.g. cancer),  
 CC immunodeficiency disorders (e.g. AIDS), autoimmune disorders  
 CC (e.g. arthritis), neurological disorders (e.g. Alzheimer's disease),  
 CC metabolic disorders (e.g. phenylketonuria), inflammatory disorders  
 CC (e.g. asthma), cardiovascular disorders (e.g. atherosclerosis),  
 CC blood-related disorders and infectious disorders (e.g. influenza). The  
 CC polynucleotides of the invention can also be used in gene therapy.  
 CC AA022915-AA023814 represent the novel human enzyme polypeptides of the  
 CC invention.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 322 AA:

Query Match 99.7%; Score 1714; DB 22; Length 322;  
 Best Local Similarity 99.7%; Pred. No. 1.1e-146;  
 Matches 316; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 MVSAGAPALGGCLGFTSTLLASFTAILAARIPVPACGRPOOLNRYVGEDSTDS 60  
 DB 6 mvsgappalggclgftstllasftailaari pvpacgkpgqlnrvvgedstde 65  
 OY 61 WPIVSIKNGTTHCAGSLTSRWVITAHCFKDNLRPYLFSVLGAMOLGNPSRSQK 120  
 DB 66 wpi v s i k n g t t h c a g s l t s r w v i t a h c f k d n l r p y l f s v l g a m o l g n p s r s q k 125  
 OY 121 VGVAVNEPYPYISKEGACADIALYRLERSTQSERVLPICLPDASTHLPNTHCWSGW 180  
 DB 126 v g v a v n e p h p y i s k e g a c a d i a l y r l e r s t q s e r v l p i c l p d a s t h l p n t h c w s g w 185  
 OY 181 GSTIOGVPLPHPOTLQKLKVPILIDSEVCSHLYWRAGSGPITTEMDMLCAGYLEGRDACL 240  
 DB 186 g s t i o g v p l p h p o t l q k l k v p i l i d s e v c s h l y w r a g s g p i t t e m d m l c a g y l e g r d a c l g 245  
 OY 241 DSGGPLMCOVDGAWLLAGIISWEGCARRNPYIISIAHRSWEKTIYGVOLGRAG 300  
 DB 246 d s g g p l m c o v d g a w l l a g i i s w e g c a r r n p y i i s i a h r s w e k t i y g v o l g r a g 305  
 OY 301 GGALRAPSGSGAARS 317  
 DB 306 g g a l r a p s g s g a a r s 322

Search completed: August 13, 2002, 08:47:16  
 Job time: 244 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OK protein - protein search, using sw model

Run on: August 13, 2002, 08:45:47 ; Search time 17.42 Seconds  
(without alignments)  
1748.583 Million cell updates/sec

Title: US-10-040-803-7

Perfect score: 1720  
Sequence: 1 MYVSGAPPALGGGLGTFSTS.....AOGGALRAPSGGGAARS 317

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database: PIR\_71.\*

1: PIR1.\*  
2: PIR2.\*  
3: PIR3.\*  
4: PIR4.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	% Match	Length	DB	ID	Description
1	607.5	35.3	343	1	A57014	proctasin (EC 3.4.21.1)
2	554	32.2	276	2	A38654	mast cell proteina
3	546	31.7	275	2	A32410	trypsin (EC 3.4.21.1)
4	538.5	31.3	274	2	JC4171	trypsin (EC 3.4.21.1)
5	536	31.2	237	2	S68702	trypsin (EC 3.4.21.1)
6	532	30.9	274	2	A45754	trypsin (EC 3.4.21.1)
7	530.5	30.8	275	2	B35863	trypsin (EC 3.4.21.1)
8	529.5	30.8	275	2	A35863	trypsin (EC 3.4.21.1)
9	522.5	30.4	275	2	C35863	trypsin (EC 3.4.21.1)
10	522	30.3	270	2	S56160	mast cell trypsin
11	518	30.1	273	2	A47246	trypsin (EC 3.4.21.1)
12	481	28.0	638	1	KOHUP	plasma kallikrein
13	475.5	27.6	625	1	KHUI	coagulation factor
14	473	27.3	638	1	KOMSP	plasma kallikrein
15	469.5	27.3	638	1	KOMSP	plasma kallikrein
16	460.5	26.8	812	1	PLBO	plasma (EC 3.4.21.1)
17	457.5	26.6	1034	1	A53663	enteropeptidase (E
18	452	26.3	417	1	S00845	hepsin (EC 3.4.21.1)
19	450	26.2	416	1	S33777	hepsin (EC 3.4.21.1)
20	450	26.2	810	2	B30848	plasma (EC 3.4.21.1)
21	446	25.9	455	2	A61545	plasma (EC 3.4.21.1)
22	446	25.9	790	1	PLPG	plasma (EC 3.4.21.1)
23	446	25.9	810	1	PLHU	polypeptide - Afri
24	445.5	25.9	1524	2	T30337	serine proteinase
25	445	25.9	786	1	A47547	enteropeptidase (E
26	444	25.8	1019	1	A56318	enteropeptidase (E
27	443.5	25.8	1035	1	A45090	apolipoprotein(a)
28	442	25.7	1420	2	A32869	plasma (EC 3.4.21.1)
29	441	25.6	460	2	B61545	plasma (EC 3.4.21.1)

30	439.5	25.6	812	1	PLMS	plasma (EC 3.4.21.1)
31	437.5	25.4	367	2	JR0104	testicular serine
32	437.5	25.4	437	2	S18407	acrosin (EC 3.4.21.1)
33	437.5	25.4	4548	1	S00657	apoptein(a) (EC 3
34	437	25.4	263	1	KYRFB	chymotrypsin (EC 3
35	437	25.4	421	1	S11674	acrosin (EC 3.4.21.1)
36	434	25.2	265	2	T15451	hypothetical prote
37	433.5	25.2	271	2	A25528	pancreatic elastas
38	432.5	25.1	436	2	JX0172	acrosin (EC 3.4.21.1)
39	430.5	25.0	366	2	JR0105	testicular serine
40	429.5	25.0	431	2	S47538	acrosin (EC 3.4.21.1)
41	428	24.9	263	2	A21195	chymotrypsin (EC 3
42	425	24.7	810	2	I46260	plasma (EC 3.4.21.1)
43	424.5	24.7	270	2	B29934	pancreatic elastas
44	422	24.5	263	2	A31299	chymotrypsin (EC 3
45	422	24.5	415	1	A34170	acrosin (EC 3.4.21.1)

## ALIGNMENTS

## RESULT 1

A57014  
proctasin (EC 3.4.21.1) precursor - human  
C/Species: Homo sapiens (man)  
C/Date: 24-May-1996 #sequence\_revision 24-May-1996 #text\_change 18-Jun-1999  
C/Accession: A57014; A54866  
R/Yu, J.X.; Chao, L.; Chao, J.  
J. Biol. Chem. 270, 13483-13489, 1995  
J. Biol. Chem. 270, 13483-13489, 1995  
A/Title: Molecular cloning, tissue-specific expression, and cellular localization of  
A/Reference number: A57014; M01D:95286644  
A/Accession: A57014  
A/Status: translated from GB/EMBL/DBJ  
A/Molecule type: mRNA  
A/Residues: 1-343 <RES>  
A/Cross-references: GB:U41351; NID:9862304; PIDN:AAQ41759.1; PID:9862305  
A/Experimental source: prostate  
A/Note: parts of this sequence were determined by protein sequencing  
R/Yu, J.X.; Chao, L.; Chao, J.  
J. Biol. Chem. 269, 18843-18848, 1994  
A/Title: Proctasin is a novel human serine proteinase from seminal fluid. Purificatio  
A/Reference number: A54866; M01D:94308140  
A/Accession: A54866  
A/Molecule type: protein  
A/Residues: 45-64 <YUA>  
C/Genetics:  
A/Gene: GDB:PRSS8  
A/Cross-references: GDB:676446; OMIM:600823  
A/Map position: 16p11.2-16p11.2  
C/Superfamily: proctasin; trypsin homology  
C/Keywords: glycoprotein; hydrolase; serine proteinase; transmembrane protein  
F:1-32/Domain: signal sequence #status predicted <SIG>  
F:33-44/Domain: product: proctasin #status predicted <CHL>  
F:33-44/Domain: proctasin light chain #status predicted <CHL>  
F:45-343/Domain: proctasin heavy chain #status predicted <CHH>  
F:45-281/Domain: trypsin homology <TRY>  
F:37-154,70-86,168-244,201-223,234-262/disulfide bonds: #status predicted  
F:85/134,228/Active site: His, Asp, Ser #status predicted  
F:159/Binding site: carboxylate (Asn) (covalent) #status experimental

## Query Match

35.3% Score 607.5; DB 1; Length 343;  
Best local similarity 44.2%; Pred. No. 6.3e-44;  
Matches 129; Conservative 42; Mismatches 108; Indels 13; Gaps 7;

QY 10 LGGGCLGTFSTL---LIASTAILNARIYPPACGRPOUNRVVGGEDSDSEMPVIVS 66  
DB 7 LGGGCLGTFSTL---LIASTAILNARIYPPACGRPOUNRVVGGEDSDSEMPVIVS 61  
QY 67 TQKNTTHHCAGSLTSTRVITTAHOFKDNLNKPYLFVSLGAMOLGNPGSRQKGVAVV 126  
DB 62 ITYESVHWGGSIVSEQWVLSAHCPSSEHKE-AIEYKLGARHQLDYSDEAKYSTIKDI 120

QY 127 EPHVYSKKEGACADIALVRLERSIOFSEVRLPTCLPDASTIHLPNTNHCWISGWSIODG 186  
Db 121 IPHSY-LOEGSGDIALQLSRPTFSRYRIPCLPAAANSFNGHCHVYTGWHAPS 179  
QY 187 VPLRHPOLQKLPPIIDSEVCSHLVWRGA-GGCP--ITEDMCAAGYLEBERDACLDSG 243  
Db 180 VSLTPRLOOLEVPLISRETNCNLVNDKPEEPHEVOEDMWCAGYEGGKACQGDSDG 239  
QY 244 GPLMCOVDAWLAGIISWEGCAERNRPGVYISLSAHSRWEKTYOGLR 295  
Db 240 GPLSCPHEGLMYLGLIYSWGDGACARRRPGVYITLASSYASWISKVTELOPR 291

## RESULT 2

A38654  
mast cell proteinase 6 (EC 3.4.21.-) precursor - mouse  
C:Species: Mus musculus (house mouse)  
C>Date: 21-Feb-1992 #sequence\_revision 17-Feb-1994 #text\_change 22-Jun-1999  
Accession: A38654; B38654; D35646; I59478  
Mol. Chem. 266, 3847-3853, 1991  
A:Title: Cloning of the cDNA and gene of mouse mast cell protease-6. Transcription by p  
A:Reference number: A38654; MUID:91139682  
A:Accession: A38654  
A:Molecule type: DNA  
A:Residues: 1-276 <REV>  
A:Cross-references: GB:M57625; NID:9200506; PIDN:AAA39987.1; PID:9200507  
A:Note: The authors translated the codon CCG for residue 24 as Ala, GAG for residue 37  
s Gly, GAG for residue 148 as Gly, GAG for residue 168 as Gly, and GAA for 185 as Gly  
A:Accession: B38654  
A:Molecule type: mRNA  
A:Residues: 1-276 <REV>  
A:Cross-references: GB:M57626; NID:9200508; PIDN:AAA39988.1; PID:9200509  
R:Reynolds, D.S.; Stevens, R.L.; Lane, W.S.; Carr, M.H.; Austen, K.F.; Serafin, W.E.  
Proc. Natl. Acad. Sci. U.S.A. 87, 3230-3234, 1990  
A:Title: Different mouse mast cell populations express various combinations of at least  
A:Reference number: A35646; MUID:90222202  
A:Accession: D35646  
A:Molecule type: protein  
A:Residues: 32-54 <RES>  
R:Huang, J.; Abirink, M.; Gobli, A.E.; Nilsson, G.; Aveskog, M.; Larsson, L.G.; Nilsson,  
Scand. J. Immunol. 38, 359-367, 1993  
A:Title: Expression of a mast cell tryptase in the human monocytic cell lines U-937 and  
A:Reference number: I59478; MUID:94023807  
A:Accession: I59478  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-276 <RES>  
A:Cross-references: GB:L31853; NID:9473480; PIDN:AAA39725.1; PID:9473481  
A:Title: Mast cell tryptase  
A:Accession: MMCP-6  
A:Introns: 24/1; 79/2; 168/1; 222/3  
C:Superfamily: trypsin; trypsin homology  
C:Keywords: hydrolase; serine proteinase; zymogen  
F:1-21/Domain: signal sequence #status predicted <SIG>  
F:22-31/Domain: activation peptide #status predicted <ACT>  
F:32-276/Product: mast cell proteinase 6 #status experimental <MAT>  
F:32-268/Domain: trypsin homology <TRY>  
F:75,122,225/Active site: His, Asp, Ser #status predicted

Query Match 32.2%; Score 554; DB 2; Length 276;  
Best Local Similarity 41.6%; Pred. No. 1.7e-39;  
Matches 116; Conservative 48; Mismatches 93; Indels 22; Gaps 9;

QY 21 LLLSLSTATINARIPVP-PACGKPOULNRVYGEDSTDEMPVIVSIO---KNTGHCA 76  
Db 6 LLLSLSTATINARIPVP-PACGKPOULNRVYGEDSTDEMPVIVSIO---KNTGHCA 76  
QY 77 GLLSLSRWVITAAHCFKDLNRPVYLSVILGAMQI--GNPGRSQKGVAVWEPHPVYSW 134  
Db 62 GLLSLSRWVITAAHCFKDLNRPVYLSVILGAMQI--GNPGRSQKGVAVWEPHPVYSW 134

QY 135 KEGACADIALVRLERSIOFSEVRLPTCLPDASTIHLPNTNHCWISGWSIODG 194  
Db 117 AEGG-ADVALLLELVVNVSTHHPISLPASSETPEPGSCWVTGMDNDDEPLPP 175  
QY 195 LOKLVPPIIDSEVCSHLVWRGA-GGCP--ITEDMCAAGYLEBERDACLDSG 250  
Db 176 LKQVKEPIVENSICDRKHYTGTDGDPPIVHDMILCAG--NTRRDSGSGSGPLVCKV 233  
QY 251 DQAMLAGIISWEGCAERNRPGVYISLSAHSRWEKTY 289  
Db 234 KGTWLOAGVYVSWEGCAOPNKGITRYVYLDWTHRY 272

## RESULT 3

A32410  
tryptase (EC 3.4.21.59) precursor - dog  
C:Species: Canis lupus familiaris (dog)  
C>Date: 12-Oct-1989 #sequence\_revision 12-Oct-1989 #text\_change 22-Jun-1999  
Accession: A32410  
R:Vanderlidge, P.; Craik, C.S.; Nadel, J.A.; Caughey, G.H.  
Biochemistry 28, 4148-4155, 1989  
A:Title: Molecular cloning of dog mast cell tryptase and a related protease: structur  
A:Reference number: A32410; MUID:89352460  
A:Accession: A32410  
A:Molecule type: mRNA  
A:Residues: 1-275 <VAN>  
A:Cross-references: GB:M24664; NID:9163982; PIDN:AAA30854.1; PID:9163983; GB:J02862  
C:Superfamily: trypsin; trypsin homology  
C:Keywords: hydrolase; serine proteinase; zymogen  
F:1-21/Domain: signal sequence #status predicted <SIG>  
F:22-30/Domain: activation peptide #status predicted <ACT>  
F:31-275/Product: tryptase #status predicted <MAT>  
F:31-267/Domain: trypsin homology <TRY>  
F:74,121,224/Active site: His, Asp, Ser #status predicted

Query Match 31.7%; Score 546; DB 2; Length 275;  
Best Local Similarity 40.9%; Pred. No. 8e-39;  
Matches 112; Conservative 46; Mismatches 102; Indels 14; Gaps 6;

QY 23 LLASRAILNARIPVP-PACGKPOULNRVYGEDSTDEMPVIVSIO---KNTGHCA 79  
Db 5 LVLLALALLGSL-VVPSPAPGALQVGVGREGAPGSKPQVSLRLKQYWRIGGSL 63  
QY 80 LTRSRVITAAHCFKDLNRPVYLSVILGAMQI--GNPGRSQKGVAVWEPHPVYSK 139  
Db 64 IHPQWVITAAHCFKDLNRPVYLSVILGAMQI--GNPGRSQKGVAVWEPHPVYSK 139  
QY 140 ADIALVRLERSIOFSEVRLPTCLPDASTIHLPNTNHCWISGWSIODG 199  
Db 120 ADIALVRLERSIOFSEVRLPTCLPDASTIHLPNTNHCWISGWSIODG 199  
QY 200 VPIIDSEVCSHLVWRGA-GGCP--ITEDMCAAGYLEBERDACLDSG 250  
Db 180 VPIIDSEVCSHLVWRGA-GGCP--ITEDMCAAGYLEBERDACLDSG 250  
QY 256 LAGIISWEGCAERNRPGVYISLSAHSRWEKTY 289  
Db 238 QAGVYVSWEGCAOPNKGITRYVYLDWTHRY 272

## RESULT 4

JC4171  
tryptase (EC 3.4.21.59) precursor - rat  
N:Alternate names: mast cell tryptase  
C:Species: Rattus norvegicus (Norway rat)  
C>Date: 27-Aug-1995 #sequence\_revision 27-Oct-1995 #text\_change 20-Jun-2000  
Accession: JC4171  
R:Ido, H.; Itoh, H.; Tomita, M.; Murakumo, Y.; Kobayashi, T.; Maruyama, H.; Osada, Y.  
J. Biochem. 118, 210-215, 1995  
A:Title: cDNA sequencing and expression of rat mast cell tryptase.  
A:Reference number: JC4171; MUID:96015171



## RESULT 7

B35863

tryptase (EC 3.4.21.59) II precursor - human  
N:Alternate names: tryptase beta

C:Species: Homo sapiens (man)

C:Date: 03-Feb-1994 #sequence, revision 03-Feb-1994 #text\_change 21-Jul-2000

C:Accession: B35863; A37193; I59473

R:VanderSluis, P.; Ballinger, S.M.; Tam, E.K.; Goldstein, S.M.; Craik, C.S.; Caughey, G.

Proc. Natl. Acad. Sci. U.S.A. 87, 3811-3815, 1990

A:Title: Human mast cell tryptase: multiple cDNAs and genes reveal a multigene serine pr

A:Reference number: A35863; MUID:90251647

A:Accession: B35863

A:Molecule type: mRNA

A:Residues: 1-275 &lt;VAN&gt;

A:Cross-references: GB:M33492; NID:9339982; PIDN:AAA36779.1; PID:9339983

A:Note: residues 2-275 are derived from mRNA; residue one was inferred from the genomic

R:Miller, J.S.; Moxley, G.; Schwartz, L.B.

J. Clin. Invest. 86, 864-870, 1990

A:Title: Cloning and characterization of a second complementary DNA for human tryptase.

A:Reference number: A37193; MUID:90369005

A:Accession: A37193

A:Molecule type: mRNA

A:Residues: 1-275 &lt;MIT&gt;

A:Cross-references: GB:M37488; NID:9179583; PIDN:AAA51843.1; PID:9179584

R:Blom, T.; Hellman, L.

Scand. J. Immunol. 37, 203-208, 1993

A:Title: Characterization of a tryptase mRNA expressed in the human basophil cell line K

A:Reference number: I59473; MUID:93166209

A:Accession: I59473

A:Molecule type: translated from GB/EMBL/DBJ

A:Residues: 1-275 &lt;RES&gt;

A:Cross-references: GB:S5551; NID:9265666; PIDN:AAD13876.1; PID:94261576

A:Experimental source: basophil cell line KUB12

C:Genetics:

A:Gene: GDB:TPS1

A:Cross-references: GDB:125890; OMIM:191080

A:Map position: 16pter-16qter

C:Superfamily: trypsin; trypsin homology

C:Keywords: hydrolase; serine proteinase; zymogen

F:1-21/Domain: signal sequence #status predicted &lt;SIG&gt;

F:22-30/Domain: activation peptide #status predicted &lt;ACT&gt;

F:31-275/Product: trypsin I #status predicted &lt;MAT&gt;

F:31-267/Domain: trypsin homology &lt;TRY&gt;

F:74,121,224/Active site: His, Asp, Ser #status predicted

## Query Match

30.8%; Score 530.5; DB 2; Length 275;  
Best Local Similarity 39.2%; Pred. No. 1.6e-37;

Matches 113; Conservative 53; Mismatches 85; Indels 37; Gaps 9;

```

OY 20 SLLSLASTALINMARIPVPACGKPOQLRVGSDSTSEMPVIVSIQKNG---THHCA 76
   :||||| : | | | | | : | | | | | : | | | | | : | | | | | : |
Db 3 NLLSLALPVL--ASRAVAAPAPGALORVIGVGEAPRSKMPQVSLRVHGPTVMHFCG 60
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
OY 77 GSLLSRWVITTAHC-----FKD-----NLNKPYLESVLLGAMOLGNGSRQKGVAVY 126
   ||| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 61 GSLHPQWVLTAAHCVGPDVKDLALRVOLRQHLV-----YQDQLLPVSRI 107
   ||| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
OY 127 EPRP--VYSMKEGACADIALVRLRSIOFSERVLPICLPDASIHLPNTNHCISGKSTOD 185
   || :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 108 IVHPQFYTAQIG--ADIALLELEPEPVKSSVHVTLPAPASETPPGMPGCVWTGMDVDN 165
   || :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
OY 186 GVPLRPHTLOKLVPIIDSEVCSHLVYWGAGGCP---ITEDMLCAGYLEGERDACIGD 241
   || :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 166 DERLPFRPPLKQVKVIMENHICDAKYHLGAYTGDDVRIYRDMILCAG--NTRRDSQCGD 223
   || :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
OY 242 SGGPLMCQVDGAWLLAGTISWEGCAERNRPQVYISLSAHRSWVEKTV 289
   ||| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 224 SGGPLVCVNGTWLDAGVSWGEGCAQPNRPQVYISLTVYLYDMIHNYV 271

```

## RESULT 8

A35863

tryptase (EC 3.4.21.59) I precursor - human  
C:Species: Homo sapiens (man)

C:Date: 12-Oct-1990 #sequence, revision 12-Oct-1990 #text\_change 21-Jul-2000

C:Accession: A35863; D35863; A60939; A39326

R:VanderSluis, P.; Ballinger, S.M.; Tam, E.K.; Goldstein, S.M.; Craik, C.S.; Caughey, G.

Proc. Natl. Acad. Sci. U.S.A. 87, 3811-3815, 1990

A:Title: Human mast cell tryptase: multiple cDNAs and genes reveal a multigene serine

A:Reference number: A35863; MUID:90251647

A:Accession: A35863

A:Molecule type: DNA

A:Residues: 1-275 &lt;VAN&gt;

A:Cross-references: GB:M33494; NID:93927804; PIDN:AAC83172.1; PID:9339977

A:Accession: D35863

A:Molecule type: mRNA

A:Residues: 1-275 &lt;VA2&gt;

A:Cross-references: GB:M33491

R:Butterfield, J.H.; Weller, D.A.; Hunt, L.W.; Wynn, S.R.; Roche, P.C.

J. Leukoc. Biol. 47, 409-419, 1990

A:Title: Purification of tryptase from a human mast cell line.

A:Reference number: A60939; MUID:90244210

A:Accession: A60939

A:Molecule type: protein

A:Residues: 31-38 &lt;CNO&gt;

A:Experimental source: mast cell

A:Note: 44-Gly was also found

R:Cromlish, J.A.; Seidah, N.G.; Marcinkiewicz, M.; Hamelin, J.; Johnson, D.A.; Chretien

J. Biol. Chem. 262, 1363-1373, 1987

A:Title: Human pituitary trypsin: molecular forms, NH-2-terminal sequence, immunocyt

A:Reference number: A39326; MUID:87109258

A:Accession: A39326

A:Molecule type: protein

A:Residues: 31-38 &lt;CNO&gt;

A:Experimental source: pituitary

C:Genetics:

A:Intons: 21/1; 78/2; 177/1; 221/3

C:Superfamily: trypsin; trypsin homology

C:Keywords: hydrolase; serine proteinase; zymogen

F:1-21/Domain: signal sequence #status predicted &lt;SIG&gt;

F:22-30/Domain: activation peptide #status predicted &lt;ACT&gt;

F:31-275/Product: trypsin I #status experimental &lt;MAT&gt;

F:31-267/Domain: trypsin homology &lt;TRY&gt;

F:74,121,224/Active site: His, Asp, Ser #status predicted

## Query Match

30.8%; Score 529.5; DB 2; Length 275;  
Best Local Similarity 39.2%; Pred. No. 2e-37;

Matches 113; Conservative 52; Mismatches 86; Indels 37; Gaps 9;

```

OY 20 SLLSLASTALINMARIPVPACGKPOQLRVGSDSTSEMPVIVSIQKNG---THHCA 76
   :||||| : | | | | | : | | | | | : | | | | | : | | | | | : |
Db 3 NLLSLALPVL--ASRAVAAPAPGALORVIGVGEAPRSKMPQVSLRVHGPTVMHFCG 60
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
OY 77 GSLLSRWVITTAHC-----FKD-----NLNKPYLESVLLGAMOLGNGSRQKGVAVY 126
   ||| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 61 GSLHPQWVLTAAHCVGPDVKDLALRVOLRQHLV-----YQDQLLPVSRI 107
   ||| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
OY 127 EPRP--VYSMKEGACADIALVRLRSIOFSERVLPICLPDASIHLPNTNHCISGKSTOD 185
   || :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 108 IVHPQFYTAQIG--ADIALLELEPEPVKSSVHVTLPAPASETPPGMPGCVWTGMDVDN 165
   || :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
OY 186 GVPLRPHTLOKLVPIIDSEVCSHLVYWGAGGCP---ITEDMLCAGYLEGERDACIGD 241
   || :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 166 DERLPFRPPLKQVKVIMENHICDAKYHLGAYTGDDVRIYRDMILCAG--NTRRDSQCGD 223
   || :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
OY 242 SGGPLMCQVDGAWLLAGTISWEGCAERNRPQVYISLSAHRSWVEKTV 289
   ||| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 224 SGGPLVCVNGTWLDAGVSWGEGCAQPNRPQVYISLTVYLYDMIHNYV 271

```

RESULT 9

C35863

```

tryptase (EC3.4.21.59) III precursor - human
C:Species: Homo sapiens (man)
C:Date: 03-Feb-1994 #sequence_revision 03-Feb-1994 #text_change 15-Jun-2001
C:Accession: G35863; A35863; A38893
R:VanderSlite, P.; Ballinger, S.M.; Tam, E.K.; Goldstein, S.M.; Craik, C.S.; Caughey, G.
Proc. Natl. Acad. Sci. U.S.A. 87, 3811-3815, 1990
A:Title: Human mast cell tryptase: multiple cDNAs and genes reveal a multigene serine pr
A:Reference number: A35863; MUID:90251647
A:Accession: G35863
A:Molecule type: mRNA
A:Residues: 9-275 <VAN>
A:Accession: A35863
A:Molecule type: DNA
A:Residues: 1-9 <VA2>
A:Cross-references: GB:M3494; NID:g3927804; PIDN:AAC83172.1; PID:g339977
A>Note: the first nine residues of this sequence are inferred from genomic DNA of trypt
R:VanderSlite, P.
Reference number: A38893
Accession: A38893
Molecule type: mRNA
Residues: 9-131,'K',132-275 <VA3>
A:Cross-references: GB:M33493; NID:g339984; PIDN:AAA36780.1; PID:g339985
C:Superfamily: trypsin; trypsin homology
C:Keywords: hydrolase; serine proteinase; zymogen
F:1-21/Domain: signal sequence #status predicted <SIG>
F:22-30/Domain: activation peptide #status predicted <ACT>
F:31-275/Product: tryptase I #status predicted <MAT>
F:31-267/Domain: trypsin homology <TRY>
F:74,121,224/Active site: His, Asp, Ser #status predicted

Query Match 30.4%; Score 522.5; DB 2; Length 275;
Best Local Similarity 38.9%; Pred. No. 7.8e-37;
Matches 112; Conservative 52; Mismatches 87; Indels 37; Gaps 9;

QY 20 SLLLSALLNARIYPPACGKPOLNRYVGEDSTSEMPWIVSIQ--KNGTHICA 76
DB 3 NLLLLAPLV--ASRAYAAPAPGOALQGVIGGGEAPRSKPMQVSLRFRDYMWHFCG 60
QY 77 GSLITSRVNTIAAC-----FRD-----NLNKPYLEFVLLGWMQGNRSGRSQKYGVMV 126
DB 61 GSLIHQWVLTAAHCVPDQKDLAALRVQLREQHLY-----YQDLLPVSRI 107
QY 127 EHPH-VYSMEGACADIALVRLERSIOFSERVLPICLPDASIHLPNTHCWISGWSIOD 185
DB 108 IYHPOFTYTAIG--ADTALLELEPVNVSSHVHTVTLPPASSETPPGMPQWVTGMDVDN 165
QY 186 GVLPHPOFTLOKLPFIIDSEVCSHLTWKRGAGOGP---ITEDMCAAGLEBERACTACD 241
DB 166 DEHLPPFPFLKQKRVPLMEHNDICAKTHLGAVTGDDVRIYVDMWICAG--NTRRDSOGD 223
QY 242 SGGPLMCQVDGAWMLLAGIISWEGECARNRPGVYISLSAHRSWEKIV 289
DB 224 SGGPLCKVKGVTWLGAGVWSMGECAQPNRPGIVRTYTYLDWZHHYV 271

RESULT 10
56160
mast cell tryptase precursor - Mongolian jird
C:Species: Meriones unguiculatus (Mongolian jird)
C:Date: 27-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 20-Jun-2000
C:Accession: S56160
R:Butchko, Y.; Ide, H.; Itoh, H.; Tomita, M.; Kobayashi, T.; Maruyama, H.; Horii, Y.; N
Biochem. J. 309, 921-926, 1995
A:Title: Cloning of the cDNA encoding mast cell tryptase of Mongolian gerbil, Meriones u
A:Reference number: S56160; MUID:95366971
A:Accession: S56160
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-270 <MDR>
A:Cross-references: EMBL:D31789; NID:g517122; PIDN:BAA06598.1; PID:g517123
C:Superfamily: trypsin; trypsin homology

```

```

F:26-262/Domain: trypsin homology <TRY>

Query Match      30.3%: Score 522; DB 2; Length 270;
Best Local Similarity 38.0%: Pred. No. 8.5e-37;
Matches 109; Conservative 52; Mismatches 84; Indels 42; Gaps 9;

QY 21 LLLLLSTALLNARIPVPACGKPOOLNNVGGEDSTOSEMWIVSIOKNGT---HHCAG 77
    ||||| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 4 LLLLLLPFLSLMHRSPFLQEWG-----IYGGGLAPGKMKPMVOYSLRNETIYMHKFCGG 56
    ||||| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |

QY 78 SLTTSRWVITTAHC-----FKDNLNKPFLFVLLGAMOLNPGSRQKGVAWVE 127
    || : : : : ||||| : : : : : | : : : : : | : : : : : | : : : : : |
Db 57 SLIHQWVWLTAAHCQPTTADPNKRVOLRKQYL-----YHDLILAVSRIL 103
    || : : : : ||||| : : : : : | : : : : : | : : : : : | : : : : : |

QY 128 PHPV-YSMKEGACADIALVRLERSIOFSERVPICLPASLHPNTHCMTISGMSITDGG 186
    || : : : : ||||| : : : : : | : : : : : | : : : : : | : : : : : |
Db 104 THPTTYAIVQNG--ADIALLELKNPYNISHPVSLPASETFPSGTLCWVIGMGIIND 161
    || : : : : ||||| : : : : : | : : : : : | : : : : : | : : : : : |

QY 187 VPLPHPQTLQIKVPIIDSEVCSHLYMRGAGQGP---TTEMLCAGYLEGERACLSDS 242
    || : : : : ||||| : : : : : | : : : : : | : : : : : | : : : : : |
Db 162 VSLPPPEPLKEOVQVPEVENQDLTKYHHGQVYTGNDIHTVRDMLCAGN-EG-HDSCQGDS 219
    || : : : : ||||| : : : : : | : : : : : | : : : : : | : : : : : |

QY 243 GGLPLMCQVDDGAMVLAGIITISWEGCEAKRRNPPGVYISLSAHSRNVKRTIY 289
    ||||| : : : : ||||| : : : : : | : : : : : | : : : : : | : : : : : |
Db 220 GGPLVCKVNGTINLQAGVYVSWEGGCLPMPRRGIIYTRVITYIYLDIMIRHY 266
    ||||| : : : : ||||| : : : : : | : : : : : | : : : : : | : : : : : |

```

```
Search completed: August 13, 2002, 08:48:00
Job time: 133 sec
```

Tue Aug 13 16:27:47 2002

us-10-040-803-7.rpx

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 13, 2002, 08:47:37 ; Search time 13.48 Seconds

(without alignments)  
910.541 Million cell updates/sec

Title: US-10-040-803-7

Perfect score: 1720  
Sequence: 1 MVVSGAPPALGGCLGFTFS.....AQGGALRAPSGSGAARS 317

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Sched: 105224 seqs, 38719550 residues  
number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database: SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1720	100.0	317	1 BSS4_HUMAN	O9gzr4 homo sapien
2	1270.5	73.9	306	1 BSS4_MOUSE	O9er10 mus musculu
3	739	43.0	290	1 MPN_HUMAN	O9bqr3 homo sapien
4	607.5	35.3	343	1 PSS8_HUMAN	O16651 homo sapien
5	584	34.0	342	1 PSS8_RAT	O9es87 rattus norv
6	578.5	33.6	273	1 TRYT_SHEEP	O9xsm2 ovis aries
7	574.5	33.4	342	1 PSS8_MOUSE	O9esd1 mus musculu
8	569	33.1	321	1 TRYG_HUMAN	O9nrr2 homo sapien
9	554	32.2	276	1 MCT6_MOUSE	P21845 mus musculu
10	546	31.7	275	1 TRYT_CANFA	P15944 canis fami
11	538.5	31.3	274	1 MCT6_RAT	P50343 rattus norv
12	532.5	31.0	275	1 TRYT_PIG	O9n2d1 sus scrofa
13	531.5	30.9	314	1 TEST_HUMAN	O9yem0 homo sapien
14	530.5	30.8	275	1 TRB2_HUMAN	P20231 homo sapien
15	530	30.8	324	1 TEST_MOUSE	O9jhh7 mus musculu
16	529.5	30.8	275	1 TRB1_HUMAN	O15661 homo sapien
17	526.5	30.6	275	1 TRYA_HUMAN	P15157 homo sapien
18	522	30.3	270	1 TRYT_MERON	P50342 metiones un
19	519.5	30.2	273	1 MCT7_RAT	P27435 rattus norv
20	519.5	30.2	311	1 TRYG_MOUSE	O9qul7 mus musculu
21	518	30.1	273	1 MCT7_MOUSE	O02844 mus musculu
22	490.5	28.5	455	1 TMS5_MOUSE	O9er04 mus musculu
23	482.5	28.1	457	1 TMS5_HUMAN	O9hs83 homo sapien
24	481	28.0	638	1 KAL_HUMAN	P03952 homo sapien
25	475.5	27.6	638	1 KAL_MOUSE	P03951 homo sapien
26	473	27.5	638	1 KAL_MOUSE	P26262 mus musculu
27	472.5	27.5	437	1 TMS4_HUMAN	O9nrs4 homo sapien
28	472.5	27.5	454	1 TMS4_HUMAN	P57727 homo sapien
29	469.5	27.3	638	1 KAL_RAT	P14772 rattus norv
30	463	26.9	416	1 HEP8_MOUSE	O35453 mus musculu
31	462	26.9	492	1 TMS2_HUMAN	O15393 homo sapien
32	460.5	26.8	812	1 PLAN_BOVIN	P06868 bos taurus
33	457.5	26.6	1034	1 ENTK_PIG	P98074 sus scrofa

34	453.5	26.4	490	1 TMS2_MOUSE	O9jig8 mus musculu
35	452	26.3	417	1 HEP8_HUMAN	P05981 homo sapien
36	451.5	26.2	855	1 STR1_MOUSE	P56677 mus musculu
37	450	26.2	416	1 HEP8_RAT	O05511 rattus norv
38	450	26.2	810	1 PLAN_MACAU	P12545 macaca mula
39	446	25.9	338	1 PLAN_HORSE	P80010 equus cabal
40	446	25.9	790	1 PLAN_PIG	P06867 sus scrofa
41	446	25.9	810	1 PLAN_HUMAN	P00747 homo sapien
42	445	25.9	333	1 PLAN_CANFA	P80009 canis fami
43	445	25.9	786	1 STUB_DROME	O05319 drosophila
44	444	25.8	1019	1 ENTK_HUMAN	P98073 homo sapien
45	443.5	25.8	1035	1 ENTK_BOVIN	P98072 bos taurus

## ALIGNMENTS

```

RESULT 1
ID      BSS4_HUMAN          STANDARD:      PRT:      317 AA.
AC      O9GZNR4: 043342:
DT      16-OCT-2001 (Rel. 40, Created)
DT      16-OCT-2001 (Rel. 40, Last sequence update)
DT      16-OCT-2001 (Rel. 40, Last annotation update)
DE      Brain-specific serine protease 4 precursor (EC 3.4.21.-) (BSSP-4)
DE      (SP0011A).
GN      PRSS26 OR PRSS22 OR BSSP4.
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX      NCBI_TaxID=9606;
[1]
RP      SEQUENCE FROM N.A.
RC      TISSUE=Brain;
RA      Mitsui S., Okui A., Komimami K., Yamaguchi N.;
RT      "Cloning and characterization of a human brain-specific serine
RT      protease, hbssp-4."
RL      Submitted (JAN-1998) to the EMBL/Genbank/DBJ databases.
[2]
RP      SEQUENCE FROM N.A.
RC      TISSUE=Pancreas;
RA      Wong G.W., Stevens R.L.;
RT      "Identification of a new member of the chromosome 16 family of serine
RT      proteases."
RL      Submitted (NOV-2000) to the EMBL/Genbank/DBJ databases.
[3]
RP      SEQUENCE OF 47-317 FROM N.A.
RA      Ricke D.O., Bruce D., Mundt M., Doggett N., Munk C., Saunders E.,
RA      Robinson D., Jones M., Buckingham J., Chasteen L., Thompson S.,
RA      Goodwin L., Bryant J., Tesmer J., Meincke L., Longmire J., White S.,
RA      Deng S., Tatum O., Campbell C., Fawcett J., Deaven L.;
RL      Submitted (DEC-1997) to the EMBL/Genbank/DBJ databases.
-1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
-1- TRYPSIN FAMILY.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@sib-sib.ch).
-----
EMBL: AB010779; BAB20263.1; -
EMBL: AF321182; AAG35070.1; -
EMBL: AC003965; AAB93671.1; -
MEROPS: S01.252; -
InterPro: IPR001314; Chymotrypsin.
InterPro: IPR001254; Trypsin.
Pfam: PF00089; trypsin.1
PRINTS: PR00722; CHYMOTRYPSIN.
SMART: SM00020; TRYP_Spc: 1.
PROSITE: PS50240; TRYPSIN_DOM: 1.

```

CC	entitles requires a license agreement (see http://www.isb-sib.ch/announce					
CC	or send an email to license@isb-sib.ch).					
CC	-----sib-stub.ch)-----					
DR	EMBL:	AB010778;	BAB20262.1.	-	-	-
DR	InterPro:	IPIR001314;	Chymotrypsin.			
DR	InterPro:	IPIR001254;	trypsin.			
DR	pfam:	Pf00089;	trypsin; 1.			
DR	PRINTS:	PR00722;	CHYMOTRYPSIN.			
DR	SMART:	SMD0020;	Tryp_Spc; 1.			
DR	PROSITE:	PS00240;	TRYPSIN_DOM; 1.			
DR	PROSITE:	PS00134;	TRYPSIN_HIS; 1.			
DR	PROSITE:	PS00135;	TRYPSIN_SER; 1.			
KW	Hydrolase;	Serine protease;	Signal.			
FT	SIGNAL	1	32	POTENTIAL.		
FT	CHAIN	33	306	BRAIN-SPECIFIC SERINE PROTEASE 4.		
FT	ACT_SITE	90	90	CHARGE RELAY SYSTEM (BY SIMILARITY).		
FT	ACT_SITE	141	141	CHARGE RELAY SYSTEM (BY SIMILARITY).		
FT	ACT_SITE	242	242	CHARGE RELAY SYSTEM (BY SIMILARITY).		
FT	DISULFID	75	91	BY SIMILARITY.		
FT	DISULFID	175	248	BY SIMILARITY.		
FT	DISULFID	208	227	BY SIMILARITY.		
FT	DISULFID	238	266	BY SIMILARITY.		
FT	CARBOND	70	70	N-LINKED (GLCNAC... ) (POTENTIAL).		
SO	SEQUENCE	306 AA;	33262 MW;	FBBF03C0C285E7E8 CRC64;		
Query Match		73.9%;	Score 1270.5;	DB 1;	Length 306;	
Best Local Similarity		76.8%;	Pred. No. 1.5e-102;			
Matches	232;	Conservative	25;	Mismatches	44;	Indels 1; Gaps 1;
OY	1	MVSGAPPALGGCGGCTGTSTLLASTALLNARIIPVPAGCPPOOLNRVGGEDESTSE	60			
Db	1	MISSPPPALGGDGPDSILLVLVTAPISAATIRVSDDCKPOOLNNRVGGEDESMQAQ	60			
OY	61	WPVIYSIKNGTGHHCAGSLTSRWYTAAHCRKDMLNKPYLESVYLGAQMOLNPSPSRSQ	120			
Db	61	WPVIYSIILKNSHHCAGSLTFNWVVTAHCRKSAMDRLSFVTLGAMKLGSPPGRSQ	120			
OY	121	VGVAVVEPHVYSWKEGACADIALVRLEBSIOFSERVLPICLDPASIHLPPTHGWISGW	180			
Db	121	VGIADVLPHPHYRSWKEGTHADIALVRLESIIOSFERLIPLCLPDSSVRLPPTDCWIAGW	180			
OY	181	GSIODGVLPHPQTLOTKLVPIIDSEVCSHLYWRAGOGPIREDMIGAGYLEGERDACLG	240			
Db	181	GSIODGVLPHPQTLOTKLVPIIDSELCKSLTYWRAGOEAITEGMLCAGYLEGERDACLG	240			
OY	241	DSCGELMCQVDGAWLLLAGIISWGECGCAERNRPYYIVLSAHSRSWEKIYGQVQLRGRAQG	300			
Db	241	DSCGELMCQVDHWMLLTGLIISWGECCAD-DREGVYTSLLAHRSWQRITVOGQVLRGYLAD	299			
OY	301	GG 302				
Db	300	SG 301				
RESULT	3					
MPN_HUMAN		STANDARD:	PRT:	290	AA.	
ID	MPN_HUMAN					
AC	OPB0R3;					
DT	16-OCT-2001	(Rel. 40,	Created)			
DT	16-OCT-2001	(Rel. 40,	Last sequence update)			
DT	16-OCT-2001	(Rel. 40,	Last annotation update)			
DE	Marapsin precursor (EC 3.4.21.-).					
GN	MPN.					
OS	Homo sapiens (Human).					
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.					
OX	NCBI_Taxid=9606;					
RN	[1]					
RP	SEQUENCE FROM N.A.					
RA	Fortunato M., Dando P.M., Rawlings N.D., Barrett A.J.;					
RT	"Cloning, sequencing and expression of marapsin, a human serine					
RT	protease."					



RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE  
 CC TRYPSIN FAMILY.  
 CC -----  
 CC THE SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: AJ306593; CAC35467.1; -  
 DR MEROPS: S01.074; -  
 DR InterPro: IPR001314; Chymotrypsin.  
 DR InterPro: IPR001254; Trypsin.  
 DR Pfam: PF00089; trypsin.1.  
 DR PRINTS: PR00722; CHYMOTRYPSIN.  
 DR SMART: SM00020; TRYP\_SPC; 1.  
 DR PROSITE: PS50240; TRYP\_SIN\_DOM; 1.  
 DR PROSITE: PS00134; TRYP\_SIN\_HIS; 1.  
 DR PROSITE: PS00135; TRYP\_SIN\_SER; 1.  
 KW Hydrolyase; Serine protease; zymogen; signal; glycoprotein.  
 FT SIGNAL 1 22 POTENTIAL.  
 FT PROPEP 23 34 MARAPST.  
 FT CHAIN 35 290 SERINE PROTEASE.  
 FT DOMAIN 35 277 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT ACT\_SITE 75 75 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT ACT\_SITE 124 124 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT DISULFID 60 76 BY SIMILARITY.  
 FT DISULFID 158 235 BY SIMILARITY.  
 FT DISULFID 191 214 BY SIMILARITY.  
 FT DISULFID 225 253 BY SIMILARITY.  
 FT CARBOHYD 55 55 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 79 79 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SO SEQUENCE 290 AA; 31940 MM; 67BDC93EC70BF7B CRC64;

Query Match 43.0%; Score 739; DB 1; Length 290;  
 Best Local Similarity 48.8%; Pred. No. 1,1e-56;  
 Matches 139; Conservative 49; Mismatches 87; Indels 10; Gaps 4;

DB 10 LILL- --- FGSQRAKAAATACGRPMALRMVGGDTQGEPMVOYSIQKNSHFCGSLI 65  
 21 LILLASTALNARIPVPAGCKPOOLNRYVGGEDSTDEMPYVSIQKNGHHCAGSL 80  
 81 TSRWVITAAHCFKNDLNKPYLFSVLLGAMOLNPGSRQKGVANVEHPYVSMKEGACA 140  
 66 AEWQWLTAAHCFR-NMSELSYQVLLGAROLVQPGPHMYARVQVESNPLYQ-CTASSA 123  
 DB 141 DIALVLEHSEISOFSEKRVLPICLPDASIHLPPTHGHWISGMSIQGVLPHPQTLQKLV 200  
 124 DVALVELLEAPVPEFTNYILPVCLDPDSVIFETGMNWMVGMSPSEEDLLPERILQKLV 183  
 DB 201 PTLDSVCSHLKWRAGSG- ---PTEDMLCAGYLEGERDACLGSGLPACQVDMGL 256  
 184 PTLIDPKMLLSKOTERGYQKTKNDMLCAGFEKCKDACKGSGGPIYLVQGSMLQ 243  
 DB 257 AGLIISWGECAGRRNRPYVVISLSAHSRWEKIVQGVOLGRRAAG 301  
 244 AGVISMGECARQNRPYVIRVTAHNMVIRIRPKLQFQPARLG 288

RESULT 4  
 PSS8\_HUMAN STANDARD; PRT; 343 AA.  
 ID PSS8\_HUMAN  
 AC 01651; OSUCAS3;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Prostatin precursor (EC 3.4.21.-).  
 GN PRSS8.

OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxId=9606;  
 RN (1)  
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 RX TISSUE-Prostate;  
 RX MEDLINE=95286644; PubMed=7768952;  
 RA Yu J.X., Chao L., Chao J.;  
 RT Molecular cloning, tissue-specific expression, and cellular  
 RT localization of human prostatin mRNA.  
 RL J. Biol. Chem. 270:13483-13489(1995).  
 RN (2)  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-Placenta;  
 RA Strausberg R.;  
 RN Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.  
 RP (3)  
 RP SEQUENCE OF 45-64.  
 RC MEDLINE=94308140; PubMed=8034638;  
 RA Yu J.X., Chao L., Chao J.;  
 RT Prostatin is a novel human serine protease from seminal fluid.  
 RT Purification, tissue distribution, and localization in prostate  
 RT gland.  
 RL J. Biol. Chem. 269:18843-18848(1994).  
 CC -1- FUNCTION: POSSESSES A TRYPSIN-LIKE CLEAVAGE SPECIFICITY.  
 CC -1- SUBUNIT: HETERODIMER OF TWO CHAINS, LIGHT AND HEAVY, HELD BY A  
 CC -1- SUBCELLULAR LOCATION: MEMBRANE-BOUND. SECRETED AFTER CLEAVAGE OF  
 CC ITS C-TERMINUS.  
 CC -1- TISSUE SPECIFICITY: FOUND IN PROSTATE, LIVER, SALIVARY GLAND,  
 CC KIDNEY, LUNG, PANCREAS, COLON, BRONCHUS AND RENAL PROXIMAL TUBULAR  
 CC CELLS. IN THE PROSTATE GLAND IT MAY BE SYNTHESIZED IN EPITHELIAL  
 CC CELLS, SECRETED INTO THE DUCTS, AND EXCRETED INTO THE SEMINAL  
 CC FLUID.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE  
 CC TRYPSIN FAMILY.  
 CC -----  
 CC THE SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: L41351; AAC41759.1; -  
 DR EMBL: U33446; AAB19071.1; -  
 DR EMBL: BC001462; AAH01462.1; -  
 DR HSSP: P00763; IDPO.  
 DR MEROPS: S01.159; -  
 DR MIM: 600823; -  
 DR InterPro: IPR001314; Chymotrypsin.  
 DR InterPro: IPR001254; Trypsin.  
 DR Pfam: PF00089; trypsin.1.  
 DR PRINTS: PR00722; CHYMOTRYPSIN.  
 DR SMART: SM00020; TRYP\_SPC; 1.  
 DR PROSITE: PS50240; TRYP\_SIN\_DOM; 1.  
 DR PROSITE: PS00134; TRYP\_SIN\_HIS; 1.  
 DR PROSITE: PS00135; TRYP\_SIN\_SER; 1.  
 KW Hydrolyase; Serine protease; zymogen; signal; glycoprotein;  
 KW Transmembrane.  
 FT SIGNAL 1 29 POTENTIAL.  
 FT PROPEP 30 32 ACTIVATION PEPTIDE.  
 FT CHAIN 33 44 PROSTATIN LIGHT CHAIN.  
 FT CHAIN 45 322 PROSTATIN HEAVY CHAIN.  
 FT PROPEP 323 343  
 FT TRANSMEM 320 340  
 FT DOMAIN 45 286 POTENTIAL.  
 FT DISULFID 37 154 SERINE PROTEASE.  
 FT DISULFID 70 86 INTERCHAIN (BY SIMILARITY).  
 FT DISULFID 168 244 BY SIMILARITY.

FT DISULFID 201 223 BY SIMILARITY.  
 FT DISULFID 234 262 BY SIMILARITY.  
 FT ACT\_SITE 85 85 CHARGE RELAY SYSTEM.  
 FT ACT\_SITE 134 134 CHARGE RELAY SYSTEM.  
 FT ACT\_SITE 238 238 CHARGE RELAY SYSTEM.  
 FT CARBOHYD 159 159 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 343 AA; 36431 MW; 98DD6447F5A8C1B2 CRC64;

Query Match 35.3%; Score 607.5; DB 1; Length 343;  
 Best Local Similarity 44.2%; Pred. No. 2.9e-45;  
 Matches 129; Conservative 42; Mismatches 108; Indels 13; Gaps 7;

QY 10 LGGGGLGFTSL--LAASTAILNARIIPVPPACGKPOOLNRYVGGEDSTDEMPYIS 66  
 7 LGPGQALVAAILLYLGLRSGTAGEAEP---CGVAPQA-RITGSSAVAGQWQWVS 61  
 QY 67 IQRNGTHHCGSLTSRWYITAAHCFKRDINKLYLSVLGAMQCNPGSRQKGVAVY 126  
 62 ITVEGVHVCGLSVSEQWVLSAAHCFPEHKE-AVEVKLGAHQDLSYSEDAKVTLKDI 120  
 DB 127 EPHVYSWKEGACADIALVRLERSIOFSERVLPLCPDASIHLPNTHGMISGWSIDG 186  
 121 IPRHSY-LQHGSGDIALLOLSRITFSKIRIPCLPAANASFPNGLHCTVGMGHVAPS 179  
 QY 187 VPLPHQTLQKLVPIIDSEVCSHLYWGA-GQGP--ITEDMLCAGYLEGEPDCLGDSG 243  
 180 VSLTFEPQLQLEVPILSRFTCNCLYNIDAKPEPHFQEDMVCAGYVGGKADCGDSG 239  
 QY 244 GPLMCOVDGAMLAGIISWEGCAERNRPQVYISLSAHSWKEKYQVQLR 295  
 DB 240 GPLSCPEVGLWYLGIVSWGDACGANRPGVYTLTASSVASKVTELOPR 291

RESULT 5  
 PSS8 RAT STANDARD; PRT; 342 AA.  
 AC 09ES87; 09ER01; Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Proctasin precursor (EC 3.4.21.-).  
 GN PSS8.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.  
 OX NCBI\_TaxID=10116;  
 [1]  
 SEQUENCE FROM N.A.  
 TISSUE=Kidney;  
 Adachi M., Kitamura K., Miyoshi T., Tomita K.;  
 Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.  
 RN [12]  
 RP SEQUENCE FROM N.A.  
 RA Wang C.;  
 RT "Molecular cloning and expression of rat proctasin.";  
 RU Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.  
 CC -I- FUNCTION: POSSESSES A TRYPSIN-LIKE CLEAVAGE SPECIFICITY (BY  
 SIMILARITY).  
 CC -I- SUBUNIT: HETERODIMER OF TWO CHAINS, LIGHT AND HEAVY, HELD BY A  
 DISULFIDE BOND (BY SIMILARITY).  
 CC -I- SUBCELLULAR LOCATION: MEMBRANE-BOUND. SECRETED AFTER CLEAVAGE OF  
 ITS C-TERMINUS (BY SIMILARITY).  
 CC -I- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1, ALSO KNOWN AS THE  
 TRYPSIN FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; AB017638; BAB20281.1; -  
 DR EMBL; AF202076; AAG32641.1; -  
 DR InterPro; IPR001254; trypsin.  
 DR Pfam; PF00089; trypsin; 1.  
 DR SMART; SM00020; Tryp-spec; 1.  
 DR PROSITE; PS00240; TRYPSIN\_DOM; 1.  
 KW Hydrolase; Serine protease; zymogen; Signal; Glycoprotein;  
 KW Transmembrane.

FT SIGNAL 1 29 POTENTIAL.  
 FT PROPEP 30 32 ACTIVATION PEPTIDE (BY SIMILARITY).  
 FT CHAIN 33 44 PROCTASIN LIGHT CHAIN.  
 FT CHAIN 45 322 PROCTASIN HEAVY CHAIN.  
 FT PROPEP 323 342 BY SIMILARITY.  
 FT TRANSMEM 320 340 POTENTIAL.  
 FT DOMAIN 45 286 SERINE PROTEASE.  
 FT DISULFID 37 154 INTERCHAIN (BY SIMILARITY).  
 FT DISULFID 70 86 BY SIMILARITY.  
 FT DISULFID 168 244 BY SIMILARITY.  
 FT DISULFID 201 223 BY SIMILARITY.  
 FT DISULFID 234 262 BY SIMILARITY.  
 FT ACT\_SITE 85 85 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT ACT\_SITE 134 134 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT ACT\_SITE 238 238 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 FT CARBOHYD 159 159 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CONFLICT 17 17 I -> V (IN REF. 1).  
 FT CONFLICT 292 292 A -> V (IN REF. 1).  
 SQ SEQUENCE 342 AA; 36843 MW; 5ED1AF05D9213B98 CRC64;

Query Match 34.0%; Score 584; DB 1; Length 342;  
 Best Local Similarity 43.0%; Pred. No. 3e-43;  
 Matches 125; Conservative 43; Mismatches 113; Indels 10; Gaps 7;

QY 15 LGFTSLLILASTAILNARI--PVPPACGKPOOLNRYVGGEDSTSEMPYISCKNG 71  
 9 LGQELAFILLLILGLQ-SRIGADGTEASCGAVIO-PRITGGGSAKPGQWQWVSITNG 66  
 QY 72 THHCAGSLTSRWYITAAHCFKRDINKLYLSVLGAMQCNPGSRQKGVAVVEHPV 131  
 67 VHVCGGSLVSNQWVVSAAHCFPREHKEE-YEVKLGHAHQDLSFENDIVHTVAOITSHS 125  
 DB 132 YSMKEGACADIALVRLERSIOFSERVLPLCPDASIHLPNTHGMISGWSIDGVPPLH 191  
 126 YR-EGSGDIALIRLSPTVFSYRIPCLPAANASFPNGLHCTVGMGHVAPVELQF 184  
 DB 192 POTLOKLVPIIDSEVCSHLY-WRGAQGP--ITEDMLCAGYLEGEPDCLGDSGGLMC 248  
 185 PRPLQQLLEVPLSRFTCNCLYNINAVPEEPTTIOQDMICAGYVGGKADCGDSGGLSC 244  
 QY 249 QVDSAMLAGIISWEGCAERNRPQVYISLSAHSWKEKYQVQLRGRAQ 299  
 245 PIDGLWYLAGIVSWGDACGANRPGVYTLTSTVSMWIIHHVAVELQPRAPVQ 295

RESULT 6  
 TRYT SHEEP STANDARD; PRT; 273 AA.  
 AC 09XSM2;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DE Trypsin 2 precursor (EC 3.4.21.59).  
 OS Ovis aries (Sheep).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 OC Bovidae; Caprinae; Ovis.  
 OX NCBI\_TaxID=9940;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Pemberton A.D., Malleese S.M., Huntley J.F., Mackellar A.,  
 Collie D.D., McMillan L., Scudamore C.L., Miller H.R.P.;

"DNA sequence of sheep mast cell tryptase and its immunolocalisation in lung, skin and gut in comparison with sheep mast cell proteinase-1.";  
 Submitted (OCT-1998) to the EMBL/Genbank/DBJ databases.  
 FUNCTION: TRYPTASE IS THE MAJOR NEUTRAL PROTEASE PRESENT IN MAST CELLS AND IS SECRETED UPON THE COUPLED ACTIVATION-DEGRANULATION RESPONSE OF THIS CELL TYPE.  
 CATALYTIC ACTIVITY: Preferential cleavage: Arg-|-, Lys-|-, but with more restricted specificity than trypsin.  
 SUBUNIT: HOMOTETRAMER (By similarity).  
 SUBCELLULAR LOCATION: RELEASED FROM THE SECRETORY GRANULES UPON MAST CELL ACTIVATION.  
 SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE TRYPSIN FAMILY. STRONGEST TO OTHER TRYPTASES.

THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 EMBL: Y18224; CAB41989.1; -  
 DR HSSP: P20231; 1A00.  
 DR MEROPS: S01.143; -  
 DR InterPro: IPR001314; Chymotrypsin.  
 DR InterPro: IPR001254; Trypsin.  
 DR Pfam: PF00089; trypsin; 1.  
 DR PRINTS: PR00722; CHYMOTRYPSIN.  
 DR SMART: SM0020; TRYPSIN\_DOM; 1.  
 DR PROSITE: PS00240; TRYPSIN\_DOM; 1.  
 DR PROSITE: PS00134; TRYPSIN\_HIS; 1.  
 DR PROSITE: PS00135; TRYPSIN\_SER; 1.  
 DR Hydrolyase: Serine protease; Signal; Glycoprotein; Zymogen.  
 KW SIGNAL  
 FT PROPEP 1 18  
 FT CHAIN 29 273  
 FT ACT\_SITE 72 72  
 FT ACT\_SITE 119 119  
 FT ACT\_SITE 222 222  
 FT DISULFID 57 73  
 FT DISULFID 153 228  
 FT DISULFID 186 209  
 FT DISULFID 218 246  
 FT CARBOHYD 231 231  
 FT SEQUENCE 273 AA; 30288 MW; DE9BA79218C3E67D CRC64;

Query Match 33.6%; Score 578.5; DB 1; Length 273;  
 Best Local Similarity 41.7%; Pred. No. 7; 1e-43;  
 Matches 115; Conservative 53; Mismatches 93; Indels 15; Gaps 7;

21 LLLASTAIIINARIIPVPPACGKPOQLNRVVGSDSTDEMPWIVSIQ---KNGTHHCAG 77  
 1 LHLTALLLSL--VSAAPAGQALORSGLIKREKAPGSPMPQVSLRVDOYWRHRCGG 59  
 78 SLTSLRWITTAHCKFNKLNKPYLFVSLVGLAMQNGSGRSOKKGVAMVEHPVYSKKEG 137  
 60 SLIHQWVLAHCTIGPELOPSPFVQLRQHL---YDQRLPIRVLVPHPIYAVEN 116  
 138 ACADIAVRLERSIOFSERVLPICLPDASITLPPNTHCMTSGMSISIDGVLPHPQTLQK 197  
 117 G-ADIALIQLEPYSISRHVQPTVLPRASEFPESQCVWTVGMDVDVNGRPLRPYPLKQ 175  
 198 LKVPIDSEVCSTLYWRAGAGG---PIT-EDMLCAGYLEGRDCLDSDSGPLMCQVDGA 253  
 176 VKVPIVENSVCMDKXHSLSLTDYSVPVIOEDNLCAG--DGGRDSCQSDSGGSLVCKVNGT 233  
 254 WLAGIISWGECAERNRPVYISLSAHSRVEKIV 289  
 234 WLGAGVSWMGCAKRNRPVYITNTITSLDMIHQYV 269

RESULT 7  
 PSS8\_MOUSE STANDARD; PRT; 342 AA.  
 AC Q9ESD1;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Proctasin precursor (EC 3.4.21.-) (Channel activating protease 1).  
 GN PRS8 OR CAP1.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Vuagniaux G., Vallet V., Fowler-Jaeger N., Bens M., Farman N., Courtois-Couty N., Vandewalle A., Rossier B.C., Hummel E.;  
 RT "Activation of the amiloride-sensitive sodium channel by the mouse serine protease mCAP1 expressed in a principal kidney cell line."  
 RL Submitted (SEP-1999) to the EMBL/Genbank/DBJ databases.  
 CC -1- FUNCTION: POSSESSES A TRYPSIN-LIKE CLEAVAGE SPECIFICITY (BY SIMILARITY). ACTIVATES AMILORIDE-SENSITIVE SODIUM CHANNELS.  
 CC -1- SUBUNIT: HETERODIMER OF TWO CHAINS, LIGHT AND HEAVY, HELD BY A DISULFIDE BOND (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: MEMBRANE-BOUND. SECRETED AFTER CLEAVAGE OF ITS C-TERMINUS (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE TRYPSIN FAMILY.  
 CC -1- CAUTION: REF.1 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO A FRAMESHIFT IN POSITION 339.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 EMBL: AF188613; AAG17054.1; ALT\_FRAME.  
 DR InterPro: IPR001254; Trypsin.  
 DR Pfam: PF00089; trypsin; 1.  
 DR SMART: SM0020; TRYPSIN\_DOM; 1.  
 DR PROSITE: PS00240; TRYPSIN\_DOM; 1.  
 KW Hydrolyase; Serine protease; Zymogen; Signal; Glycoprotein; Transmembrane.  
 KW SIGNAL  
 FT PROPEP 1 29  
 FT CHAIN 33 32  
 FT CHAIN 45 44  
 FT PROPEP 323 342  
 FT TRANSMEM 320 340  
 FT DOMAIN 45 286  
 FT DISULFID 37 154  
 FT DISULFID 70 86  
 FT DISULFID 168 244  
 FT DISULFID 201 223  
 FT DISULFID 234 262  
 FT ACT\_SITE 85 85  
 FT ACT\_SITE 134 134  
 FT ACT\_SITE 238 238  
 FT CARBOHYD 110 110  
 FT CARBOHYD 159 159  
 FT SEQUENCE 342 AA; 36729 MW; 0620DE88ED187D0F CRC64;

Query Match 33.4%; Score 574.5; DB 1; Length 342;  
 Best Local Similarity 42.6%; Pred. No. 2e-42;  
 Matches 124; Conservative 46; Mismatches 110; Indels 11; Gaps 8;

Db 7 LGAGLEAVTILLLLG--LLQSGIRADGTEASGAVIO--PRITGGSGANPGQWPMOVSIIT 63  
 QY 69 KNGTHGACSLTSMRWYTAHCF-KDNLNKPFLFVSLGAMOLGPNRSQKVGAVME 127  
 Db 64 YDGNHGGSSLVNKNWVAHCFPREHSEAF--EVLGAHOLDISTNDYVHYIAQII 121  
 QY 128 PHPVYSKEGACADIALVRLERSIOFSERVLPICLPDASTIHLPRNTHCWSIGMSIODGV 187  
 Db 122 THSSYR-EEGSGQDIAFIRLSSEVTFSRVIRPICLPAANASPNGLCHCTVGTGHHVAPSV 180  
 QY 188 PLHPOTLOKLPITIDSEVCSHLV-WRGAGQCP--ITEDMLCAGLIEGRDACLDSG 244  
 Db 181 SLDPRLPRLQLEVLPLISRECSCLNINAVPEEPHTIQDMLCAGYVKGKDCACQDSSG 240  
 QY 245 PLMCOVDGAMTLAGIISWEGCAERNRPGYIISLAHRSWVEKIVQGVOLR 295  
 Db 241 PLSCPMGIMYLAGIYVWGDACGAPRNPGRVYITLTSTYASIMHHVHVLQPR 291  
 QY 8  
 Db HUMAN  
 PRTG-HUMAN STANDARD; PRT; 321 AA.  
 Q9NR02: Q9NR08; Q9C015; Q9UBB2;  
 16-OCT-2001 (Rel. 40, Created)  
 16-OCT-2001 (Rel. 40, Last sequence update)  
 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Trypsin gamma precursor (EC 3.4.21.-) (Transmembrane trypsinase).  
 GN TP5G1 OR TMT.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A. (VARIANTS GAMMA-1 AND GAMMA-2).  
 RA MEDLINE-20302813; PubMed-10843716;  
 RA Caughey G.H., Raymond W.W., Blount J.L., Hau L.W., Pallaro M.,  
 RA Wolters P.J., Verghese G.M.:  
 RT "Characterization of human gamma-tryptases, novel members of the  
 RT chromosome 16p mast cell trypsin and proctasin gene families.";  
 RL J. Immunol. 164:6566-6575(2000).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE-99452974; PubMed-10521469;  
 RA Wong G.W., Tang Y., Feyfant E., Sali A., Li L., Li Y., Huang C.,  
 RA Friend D.S., Krilis S.A., Stevens R.L.:  
 RT "Identification of a new member of the trypsin family of mouse and  
 RT human mast cell proteases which possesses a novel COOH-terminal  
 RT hydrophobic extension.";  
 RL J. Biol. Chem. 274:30784-30793(1999).  
 RN [3]  
 RP SEQUENCE OF 220-321 FROM N.A.  
 RA Miltman S., Agnew W.S.:  
 RT "Organization and alternative splicing of CACNA1H.";  
 RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.  
 CC -1- SUBCELLULAR LOCATION: MEMBRANE-ANCHORED (potential).  
 CC -1- TISSUE SPECIFICITY: Expressed in many tissues.  
 CC -1- POLYMORPHISM: There are two alleles; gamma-I and gamma-II which  
 CC differ by 5 residues.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE  
 CC TRYPSIN FAMILY. TRYPSIN SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 CC EMBL, AF191031; AAF76457.1; -  
 CC EMBL, AF195508; AAF76458.1; -  
 CC EMBL, AF175759; AAF03697.1; -  
 CC EMBL, AF175522; AAF03695.1; -

DR EMBL: AF223563; AAC48852.2; -.  
 DR MEROPS: S01.028; -.  
 DR InterPro: IPR001314; Chymotrypsin.  
 DR InterPro: IPR001254; Trypsin.  
 DR Pfam: PF00089; Trypsin; 1.  
 DR PRINTS: PR00722; CHYMOTRYPSIN.  
 DR SMART: SM00020; TRYP-SPC; 1.  
 DR PROSITE: PS0240; TRYPSIN\_DOM; 1.  
 DR PROSITE: PS00134; TRYPSIN\_HIS; 1.  
 DR PROSITE: PS00135; TRYPSIN\_SER; FALSE\_NEG.  
 KW Hydrolase; Serine protease; Signal; Glycoprotein; Zymogen;  
 KW Transmembrane; Polymorphism.  
 FT SIGNAL 1 19  
 FT CHAIN 20 36  
 FT CHAIN 38 321  
 FT TRANSMEM 284 304  
 FT ACT\_SITE 78 78  
 FT ACT\_SITE 125 125  
 FT ACT\_SITE 222 222  
 FT DISULFID 26 145  
 FT DISULFID 63 79  
 FT DISULFID 159 228  
 FT DISULFID 192 210  
 FT DISULFID 218 246  
 FT CARBOHYD 85 85  
 FT VARIANT 60 60  
 FT VARIANT 126 126  
 FT VARIANT 132 132  
 FT VARIANT 204 204  
 FT VARIANT 288 288  
 FT CONFLICT 160 160  
 FT SEQUENCE 321 AA; 33827 MW; FF7B06B3C4A962D CRC64;  
 Query Match 33.1%; Score 569; DB 1; Length 321;  
 Best Local Similarity 41.1%; Pred. No. 5.6e-42;  
 Matches 123; Conservative 44; Mismatches 108; Indels 24; Gaps 8;  
 QY 15 LGFTTSLLLASTAIALNARIPVPACGKPOQLN---RVVGGEDSDSEMPWIVSIQNG 71  
 Db 3 LGAGCLILLT---AVPGVSLRTLPQCGGRPOVSAGGRIVGNAAPGAMPQASLRRLR 59  
 QY 72 THHGACSLTSMRWYTAHCFKDNLPFLFVSLGAMOLGPNRSQKVGAVMEHPV 131  
 Db 60 MHVCGSSILSPQWVLTAAHCFSGSLNSD--YVHLGELTETLSPHST---VROIIHSS 115  
 QY 132 YSMREGACADIALVRLERSIOFSERVLPICLPDASTIHLPRNTHCWSIGMSIODGVPLP 191  
 Db 116 PSQPGTSGDALVELSVPTLSRLIPVCLPEASDFGICGWTGKGTREGEPLRP 175  
 QY 192 PHTLOKLPITIDSEVCSHLVWRGAGQCPITEDMLCAGLIEGRDACLDSGGLMCOVD 251  
 Db 176 PYSLEKVASVDETERDRY-PGPGGSILOPMLCA---RGPGGACQDSDGPLYCOVA 231  
 QY 252 GAWTLAIIISWEGCAERNRPGYIISLAHRSWVEKIVQGVOLRGAQGS---GALRAP 307  
 Db 232 GAWQAGIVSMGEGCGRNPGRVYITRVPAYVNMIRRI-----TASGSESSEGYRLP 283  
 RESULT 9  
 ID MCT6\_MOUSE STANDARD; PRT; 276 AA.  
 AC P21845; Q61962;  
 DT 01-MAY-1991 (Rel. 18, Created)  
 DT 01-AUG-1991 (Rel. 19, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Mast cell protease 6 precursor (EC 3.4.21.59) (MCP-6) (Trypsinase).  
 GN MCT6.

OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 RX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=91139682; PubMed=1995638;  
 RA Reynolds D.S., Gurley D.S., Austen K.F., Serafin W.E.;  
 RT "Cloning of the CDNA and gene of mouse mast cell protease-6.  
 RT Transcription by progenitor mast cells and mast cells of the  
 RT connective tissue subclass.";  
 RL J. Biol. Chem. 266:3847-3853(1991).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX STRAIN=LEADEN X A1;  
 RX MEDLINE=94023807; PubMed=8210998;  
 RA Huang R., Abtink M., Gosh A.E., Nilsson G., Aveskog M., Larsson L.G.,  
 RA Nilsson K., Hellman L.;  
 RT "Expression of a mast cell tryptase in the human monocytic cell lines  
 RT U-937 and Mono Mac 6.";  
 RX Scand. J. Immunol. 38:359-367(1993).  
 RN [3]  
 RP SEQUENCE FROM N.A. (SHORT FORM).  
 RX STRAIN=LEADEN X A1;  
 RX MEDLINE=95048582; PubMed=7959952;  
 RA Huang R., Hellman L.T.;  
 RT "Genes for mast-cell serine protease and their molecular evolution.";  
 RL Immunogenetics 40:397-414(1994).  
 RN [4]  
 RP SEQUENCE OF 32-54.  
 RX MEDLINE=90222202; PubMed=2326280;  
 RA Reynolds D.S., Stevens R.L., Lane W.S., Carr M.H., Austen K.F.,  
 RA Serafin W.E.;  
 RT "Different mouse mast cell populations express various combinations  
 RT of at least six distinct mast cell serine proteases.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 87:3230-3234(1990).  
 CC -1- FUNCTION: TRYPTASE IS THE MAJOR NEUTRAL PROTEASE PRESENT IN MAST  
 CC CELLS, AND IS SECRETED UPON THE COUPLED ACTIVATION-DEGRANULATION  
 CC RESPONSE OF THIS CELL TYPE.  
 CC -1- CATALYTIC ACTIVITY: Preferential cleavage: Arg-|-, Lys-|-, but  
 CC with more restricted specificity than trypsin.  
 CC -1- ALTERNATIVE PRODUCTS: TWO FORMS OF MCPT6 ARE PRODUCED BY  
 CC ALTERNATIVE SPLICING OF THE SAME GENE. THE SHORT FORM IS PROBABLY  
 CC A NON FUNCTIONAL VARIANT.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE  
 CC TRYPSIN FAMILY. TRYPTASE SUBFAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC at the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: M57626; AAA39988.1;  
 DR EMBL: M57625; AAA39987.1;  
 DR EMBL: L31853; AAA39725.1;  
 DR EMBL: X78542; CAA55288.1;  
 DR PIR: A38654; A38654.  
 DR PIR: D35646; D35646.  
 DR HSSP: P20231; IAAO.  
 DR MORDS: S01.025;  
 DR MGD: MG1:96942; MCP6.  
 DR InterPro: IPR001314; Chymotrypsin.  
 DR InterPro: IPR001254; Trypsin.  
 DR Pfam: PF00089; trypsin.1.  
 DR PRINTS: PRO0722; CHYMOTRYPSIN.  
 DR SMART: SM00020; TYP\_SPC.1.  
 DR PROSITE: PS00240; TRYPSIN\_DOM.1.  
 DR PROSITE: PS00134; TRYPSIN\_HIS.1.  
 DR PROSITE: PS00135; TRYPSIN\_SER.1.  
 KW Hydrolase; Serine protease; Signal; Glycoprotein; Zymogen;

KW Alternative splicing.  
 FT SIGNAL 1 21  
 FT PROPEP 22 31  
 FT CHAIN 32 276  
 FT ACT\_SITE 75 75  
 FT ACT\_SITE 122 122  
 FT ACT\_SITE 225 225  
 FT DISULFID 60 76  
 FT DISULFID 156 231  
 FT DISULFID 189 212  
 FT DISULFID 221 249  
 FT CARBOHYD 133 133  
 FT VARSPIC 223 230  
 FT VARSPIC 231 276  
 FT SEQUENCE 276 AA; 30927 MW; 525B2C9A04A72200 CRC64;  
 S0  
 Query Match 32.2%; Score 554; DB 1; Length 276;  
 Best Local Similarity 41.6%; Pred. No. 9,4e-41;  
 Matches 116; Conservative 48; Mismatches 93; Indels 22; Gaps 9;  
 QY 21 LLLASTAIINAAIRIPV-PACKGPOOLNRYVGGEDESDSEWPIVSIQ---KNGTHHCA 76  
 DB 6 LLLMALSLASLVYSAKRA-----NORVGIVGHASESEKMPWQVSLRKLTWTHFCG 61  
 QY 77 GSLITSRWVITTAACRKNINIKPYLESVLLGAWOL--GNFSRSQRYGVAWEHPHYSM 134  
 DB 62 GSLIHPOWVLTAAHCVPHPKSPOLRFVQLREQLYGD-----QLLSINRIYVHPHYT 116  
 QY 135 KEACADIALVRLERSIOFSEVRLPICDPASTHTLPNTHCWISGSGSIQDGVPLPHPT 194  
 DB 117 AEGG-ADVALLELEFVYNNVSTHTHPISLPASTFPFGISCWGTGWDINDDEPLPPV 175  
 QY 195 LQKLVPIIDSEVCSHLWYWGAGQG--PTTED-MLCAGYLEGERDAGDSGGPIACQV 250  
 DB 176 LKQVAPVIVENSICDKRKHGTYGTGDPFIVHDMICAG--NTRRSCGDSGGPLVCV 233  
 QY 251 DGAMLLAGIISWEGCAERNRPQVYISLSAHSRWVAKIV 269  
 DB 234 KGIWLAGVSWEGCAQPKPQITRYVYIDMIRHYV 272  
 RESULT 10  
 TRYP CANFA  
 ID TRYP CANFA STANDARD; PRT; 275 AA.  
 AC P15944;  
 DT 01-APR-1990 (Rel. 14, Created)  
 DT 01-APR-1990 (Rel. 14, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Trypsin precursor (EC 3.4.21.59).  
 OS Canis familiaris (dog).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.  
 RX NCBI\_TaxID=9615;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=89352460; PubMed=2504277;  
 RA Vanderslice P., Craik C.S., Nadel J.A., Caughey G.H.;  
 RT "Molecular cloning of dog mast cell tryptase and a related protease;  
 RT structural evidence of a unique mode of serine protease activation.";  
 RL Biochemistry 28:4148-4155(1989).  
 CC -1- FUNCTION: TRYPTASE IS THE MAJOR NEUTRAL PROTEASE PRESENT IN MAST  
 CC CELLS AND IS SECRETED UPON THE COUPLED ACTIVATION-DEGRANULATION  
 CC RESPONSE OF THIS CELL TYPE.  
 CC -1- CATALYTIC ACTIVITY: Preferential cleavage: Arg-|-, Lys-|-, but  
 CC with more restricted specificity than trypsin.  
 CC -1- SUBUNIT: HOMOTETRAMER.  
 CC -1- SUBCELLULAR LOCATION: RELEASED FROM THE SECRETORY GRANULES UPON  
 CC MAST CELL ACTIVATION.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE  
 CC TRYPSIN FAMILY. STRONGEST TO OTHER TRYPTASES.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 13, 2002, 08:47:22 ; Search time 30.21 Seconds  
(without alignments)  
1815.273 Million cell updates/sec

Title: US-10-040-803-7

Perfect score: 1720

Sequence: 1 MVSGAPALGGGCLGFTS.....AOGGALRAPSGSGAARS 317

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Number of hits satisfying chosen parameters: 562222

Database :

SPREMBL\_19:\*

1: sp.archaea:\*

2: sp.bacteria:\*

3: sp.fungi:\*

4: sp.human:\*

5: sp.invertebrate:\*

6: sp.mammal:\*

7: sp.mhc:\*

8: sp.organelle:\*

9: sp.phage:\*

10: sp.plant:\*

11: sp.protist:\*

12: sp.virus:\*

13: sp.unclassified:\*

14: sp.unclassified:\*

15: sp.virus:\*

16: sp.bacteriophage:\*

17: sp.archaea:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1213.5	70.6	297	11	088781	088781 rattus ratt
2	660.5	38.4	389	13	Q9PVX7	Q9PVX7 xenopus lae
3	603	35.1	317	13	Q9DGR3	Q9DGR3 xenopus lae
4	578.5	33.6	339	11	Q9JL44	Q9JL44 mus musculu
5	571	33.2	310	11	Q9JXC4	Q9JXC4 mus musculu
6	570	33.1	321	4	Q96RZ8	Q96RZ8 homo sapien
7	564	32.8	310	11	Q9QY29	Q9QY29 mus musculu
8	559.5	32.5	273	6	Q9XSM1	Q9XSM1 ovis aries
9	537	31.2	799	11	Q9DB10	Q9DB10 mus musculu
10	536	31.2	237	6	Q29464	Q29464 bos taurus
11	529.5	30.8	329	13	Q42272	Q42272 xenopus lae
12	524	30.5	273	11	Q921N4	Q921N4 mus musculu
13	517.5	30.1	275	4	Q96RZ6	Q96RZ6 homo sapien
14	491	28.5	806	6	Q18783	Q18783 macropus eu
15	485.5	28.2	279	11	Q99MS4	Q99MS4 mus musculu
16	480	27.9	282	11	Q9D413	Q9D413 mus musculu

17	480	27.9	322	11	Q920S2	Q920S2 mus musculu
18	475	27.6	624	11	Q91Y47	Q91Y47 mus musculu
19	473	27.5	624	11	Q9D4T3	Q9D4T3 mus musculu
20	472.5	27.5	405	4	Q96E86	Q96E86 homo sapien
21	469	27.3	624	6	Q95ME7	Q95ME7 oryctolagus
22	463.5	26.9	267	5	Q9BK47	Q9BK47 ludia foli
23	462.5	26.9	643	6	Q97506	Q97506 sus scrofa
24	462	26.9	492	4	Q96T73	Q96T73 homo sapien
25	455.5	26.5	812	11	Q9R0W3	Q9R0W3 rattus norv
26	452	26.3	454	6	Q46506	Q46506 papio hamad
27	450	26.2	334	6	Q46507	Q46507 papio hamad
28	445.5	25.9	812	11	Q91WJ5	Q91WJ5 mus musculu
29	445.5	25.9	1524	13	Q91674	Q91674 xenopus lae
30	445	25.9	787	5	Q9WEY6	Q9WEY6 drosophila
31	445	25.9	810	4	Q15146	Q15146 homo sapien
32	439.5	25.6	490	11	Q920K3	Q920K3 rattus norv
33	437.5	25.4	367	11	Q70169	Q70169 mus musculu
34	435.5	25.3	581	4	Q9BYE2	Q9BYE2 homo sapien
35	435	25.3	263	11	Q9DC86	Q9DC86 mus musculu
36	435	25.3	277	5	Q96899	Q96899 scolopendra
37	434	25.2	265	5	Q17800	Q17800 caenorhabdi
38	433	25.2	329	6	Q9GL10	Q9GL10 ovis aries
39	433	25.2	537	4	Q9BYE1	Q9BYE1 homo sapien
40	432.5	25.1	767	13	Q9DGR2	Q9DGR2 xenopus lae
41	430.5	25.0	366	11	Q70170	Q70170 mus musculu
42	430	25.0	263	11	Q9CR35	Q9CR35 mus musculu
43	429.5	25.0	261	13	Q9W7Q4	Q9W7Q4 paracichthy
44	429.5	25.0	279	11	Q9QZ74	Q9QZ74 rattus norv
45	428.5	24.9	249	13	Q9W6K0	Q9W6K0 notothenia

## ALIGNMENTS

RESULT	ID	PRELIMINARY	PRT	297 AA.
088781	088781			
AC	088781			
DT	01-NOV-1998 (TREMBLrel. 08, Created)			
DT	01-NOV-1998 (TREMBLrel. 08, Last sequence update)			
DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)			
DE	SERINE PROTEASE PRECURSOR (FRAGMENT).			
GN	BSP2.			
OS	Rattus rattus (Black rat).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.			
OX	NCBI_TaxID=10117;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-FISHER; TISSUE-BRAIN;			
RX	MEDLINE=98389725; PubMed=9722524;			
RA	Davies B.U., Pickard B.S., Steel M., Morris R.G., Lathe R.;			
RT	"Serine Proteases in Rodent Hippocampus.";			
RL	J. Biol. Chem. 273:23004-23011(1998).			
CC	-I- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE TRYPSIN FAMILY.			
DR	EMBL: A1005642; CA06644.1; -			
DR	HSSP: P00763; IDPO.			
DR	MEROPS: S01.252; -			
DR	InterPro: IPR001314; Chymotrypsin.			
DR	InterPro: IPR001254; Trypsin.			
DR	Pfam: PF00089; Trypsin. 1.			
DR	PRINTS: PR00722; CHYMOTRYPSIN.			
DR	SMART: SM00020; TRYP-Spe; 1.			
DR	PROSITE: PS0240; TRYPSIN_DOM; 1.			
DR	PROSITE: PS00134; TRYPSIN_HIS; UNKNOWN_1.			
DR	PROSITE: PS00135; TRYPSIN_SER; 1.			
KW	Hydrolase; Protease; Serine protease; Signal.			
FT	NON_TER	1		
FT	SIGNAL	<1	23	POTENTIAL.
FT	CHAIN	24	297	SERINE PROTEASE.
SO	SEQUENCE	297 AA;	32086 MW;	6DA25C633D6AB55 CRC64;

Query Match 70.6%; Score 1213.5; DB 11; Length 297;  
 Best Local Similarity 74.7%; Pred. No. 3e-104; Indels 9; Gaps 1;  
 Matches 222; Conservative 26; Mismatches 40; Indels 9; Gaps 1;

QY 6 APALGGGCGTFTSLLLASTAALNARIYPVPCGKPOQLNRVVGEDSTDEMPWIV 65  
 DB 5 SPGL-----TFILLPSATVSANIRGSPKPOQLNRVVGEDSDADAGMPWIV 55  
 QY 66 SIOKNGTHHCAGSLTSRWVITAAHCFKDNLPYFVLLGAMQNGNSRSQKGVAM 125  
 DB 56 SIIKNGSHHCAGSLTNRWVSAAHCFSSNMDKPSPYVLLGAMKLGNGCPRSQKGLAS 115  
 QY 126 VEPHPVYMWKGCADIALVRLERSIOFSEVLPICLPDASIHLPNTNCWISGMSIOD 185  
 DB 116 VLPHPRKSRKEGTHADIALVRLERPIQFSRIIPICLPDSSVHLPPNTNCWISGMSIOD 175  
 QY 186 GVPLPHPOTLOKLVPIIDSEVCSHLYWRGAGGPTTMDLCAYLEGERDACLGDGSGP 245  
 DB 176 GVPLPHPOTLOKLVPIIDSEVCSHLYWRGAGGPTTMDLCAYLEGERDACLGDGSGP 235  
 QY 246 LMCQVDCAMLLAGTISWGECACERNRPGVYISIAHRSWVEKIVQVQLRGAQGGG 302  
 DB 236 LMCQVDDHMLTGLTISWGECACERNRPGVYISIAHRSWVEKIVQVQLRGLADSG 292

RESULT 2  
 Q9PVX7 PRELIMINARY; PRT; 389 AA.  
 ID Q9PVX7;  
 AC Q9PVX7;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE PEPTIDASE SPECIFIC SERINE PROTEASE.  
 GN XESTIN.  
 OS Xenopus laevis (African clawed frog).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;  
 OC Xenopodinae; Xenopus.  
 OX NCBI\_TaxID=8355;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Yamada K.;  
 RT "The expression control of xestin by non-axial and planar  
 RT posteriorizing signals in Xenopus epidermis.";  
 RL Submitted (Oct-1998) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE  
 CC TRYPSIN FAMILY.  
 EMBL: AB018694; BA84941.1; -;  
 HSSP: P00763; IDPO.  
 MEROPS: S01.0PA; -;  
 DR Interpro: IPR001314; Chymotrypsin.  
 DR Interpro: IPR001254; Trypsin.  
 DR Pfam: PF00089; trypsin.1.  
 DR PRINTS: PR00722; CHYMOTRYPSIN.  
 DR SMART: SM00020; TRYP\_SPC.1.  
 DR PROSITE: PS50240; TRYPSIN\_DOM.1.  
 DR PROSITE: PS00134; TRYPSIN\_HIS; UNKNOWN\_1.  
 DR PROSITE: PS00135; TRYPSIN\_SER.1.  
 KW Hydrolyase; Protease; Serine protease.  
 SQ SEQUENCE 389 AA; 42375 MW; B31B4A2F5D1F6E3 CRC64;

Query Match 38.4%; Score 660.5; DB 13; Length 389;  
 Best Local Similarity 47.0%; Pred. No. 6.2e-53;  
 Matches 131; Conservative 42; Mismatches 95; Indels 11; Gaps 6;

QY 40 ACGKPOQLNRVVGEDSTDEMPWIVSIQNGTHHCAGSLTSRWVITAAHCFKDNLPK 99  
 DB 16 ACGVPIVSNIRVIGMDSKRGEMWQISLSYKSDSICGSLTDSWWTAAHCT-DSIDVS 74  
 QY 100 YLSEVLGAMQNGNSRSQKGVAMVEHPVYMWKGCADIALVRLERSIOFSEVLP 159

DB 75 Y-YTVYLGAYOLAPDNSTVSGVSKITKHPFOY-EGSSGDIALIELEKPVFTPIYL 132  
 QY 160 ICLPASIHLPNTHCWISGMSIODGVPLPHPOTLOKLVPIIDSEVCSHLYWRGAGOG 219  
 DB 133 ICLPSQDYQFAAGTWCWTKGNTIDEGFPLSPRTICKAEFAIIDSVCGRMYSSIGYI 192  
 QY 220 P-----ITEDMLCAGYLEGERDACLGDGSGPLMCQVDCAMLLAGTISWGECACERNRPGV 275  
 DB 193 PDFSFIOEDMVCAGYKEGRIDACQDSGGLVCVNNNNWVLOGLIVSMWGCAPERNRPGV 252  
 QY 276 ISLSAHSRWVEKIVQVQLRGAQGGGALRAPSQSGSA 314  
 DB 253 TKVQYQDMWK---TNVPLIVFSEGPVS-APSIGPSIA 287

RESULT 3  
 Q9DGR3 PRELIMINARY; PRT; 317 AA.  
 ID Q9DGR3;  
 AC Q9DGR3;  
 DT 01-MAR-2001 (TREMBLrel. 16, Created)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE EMERYONIC SERINE PROTEASE-1.  
 GN XESP-1.  
 OS Xenopus laevis (African clawed frog).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;  
 OC Xenopodinae; Xenopus.  
 OX NCBI\_TaxID=8355;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-20363741; PubMed=10903452;  
 RA Yamada K., Takabatake T., Takeshima K.;  
 RT "Isolation and characterization of three novel serine protease genes  
 RT from Xenopus laevis.";  
 RL Gene 252:209-216(2000).  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE  
 CC TRYPSIN FAMILY.  
 EMBL: AB038496; BAB08216.1; -;  
 DR HSSP: P00763; IDPO.  
 DR MEROPS: S01.048; -;  
 DR Interpro: IPR001314; Chymotrypsin.  
 DR Interpro: IPR001254; Trypsin.  
 DR Pfam: PF00089; trypsin.1.  
 DR PRINTS: PR00722; CHYMOTRYPSIN.  
 DR SMART: SM00020; TRYP\_SPC.1.  
 DR PROSITE: PS50240; TRYPSIN\_DOM.1.  
 DR PROSITE: PS00134; TRYPSIN\_HIS; UNKNOWN\_1.  
 KW Hydrolyase; Protease; Serine protease.  
 SQ SEQUENCE 317 AA; 34413 MW; EEC78A9F46D138FE CRC64;

Query Match 35.1%; Score 603; DB 13; Length 317;  
 Best Local Similarity 40.4%; Pred. No. 1e-47;  
 Matches 116; Conservative 52; Mismatches 103; Indels 16; Gaps 5;

QY 10 LGGGCGTFTSLLLASTAALNARIYPVPCGKPOQLNRVVGEDSTDEMPWIVSIQK 69  
 DB 1 MCKWLLYVTTLLFVSPHPSISNTTAAPILOGSVFSRIVGTDGAMPQVSLF 60  
 QY 70 NCTHHCAGSLTSRWVITAAHCFKDNLPYFVLLGAMQNGNSRSQKGVAMVE 127  
 DB 61 NSHICGGSIISDOWTLTATFCTI-EHPDLPSGCGVRLAYOLYKNPHEMTVKDDIYIN 119  
 QY 128 PH---PVYMWKGCADIALVRLERSIOFSEVLPICLPDASIHLPNTNCWISGMSIOD 184  
 DB 120 SEFNGP-----GTSGDIALLLKSSPIKFTYELPILCLPASVFPSSGTEDMITWGTG 173  
 QY 185 DGVLPHPOTLOKLVPIIDSEVCSHLYWRGAGGPTTMDLCAYLEGERDACLGDGSGP 240  
 DB 174 SEVPLQYPAITLOKLVVPIINRDSCKRMHINSIVSETELLIOSDQICAGYAGQKDCQG 233  
 QY 241 DSGGLMCQVDCAMLLAGTISWGECACERNRPGVYISIAHRSWVEK 287



DB 234 DSGGLVCKIQGFVWQAGIVSGERCAANRPGVYTFVPAVYTWLSE 280

RESULT 4  
ID 099L44 PRELIMINARY: PRT: 339 AA.

AC 099L44: 01-JUN-2001 (TREMBlrel. 17, Created)  
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)  
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
DE SIMILAR TO PROTEASE, SERINE, 8 (PROSTASIN).

OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;

SEQUENCE FROM N.A.  
Strausberg R.;  
Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.

SEQUENCE FROM N.A.

STRAIN=129/SVJ, AND SWISS; TISSUE=LUNG;  
Verghese G.M., Cauphey G.H.;

"Molecular cloning and characterization of mouse prostasin, a type I  
membrane-associated serine protease of the gamma-tryptase/prostasin  
gene family."

Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.  
-1 SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1, ALSO KNOWN AS THE

TRYPSIN FAMILY.

EMBL: BC003851; AA03851.1; -;  
EMBL: AF378086; AAL06320.1; -;  
EMBL: AF378085; AAL06319.1; -;  
DR HSP; P00734; IUVS.

DR MEROPS: S01.158; -;

DR InterPro: IPR001314; Chymotrypsin.  
DR InterPro: IPR001254; Trypsin.

DR Pfam: PF00089; trypsin; 1.  
DR PRINTS: PR00722; CHYMOTRYPSIN.

DR SMART: SM00020; TRYP-Spc; 1.

DR PROSITE: PS00240; TRYP-SIN\_DOM; 1.

DR PROSITE: PS00134; TRYP-SIN\_HIS; UNKNOWN\_1.

DR PROSITE: PS00135; TRYP-SIN\_SER; 1.

Hydrolase; Protease; Serine protease.  
KW SEQUENCE 339 AA; 36216 MW; BC2DE8BC057AE10 CRC64;

Query Match 33.6%; Score 578.5; DB 11; Length 339;  
Best Local Similarity 43.0%; Pred. No. 2e-45;  
Matches 125; Conservative 46; Mismatches 109; Indels 11; Gaps 8;

QY 10 LGGGCGTFTSLLLASTAALNAARIP-VPPACGRPQQLNRVYGGEDSTDEMPWIVSIQ 68  
DB 11 LGGGCGTFTSLLLASTAALNAARIP-VPPACGRPQQLNRVYGGEDSTDEMPWIVSIQ 68

DB 7 LGGGCGTFTSLLLASTAALNAARIP-VPPACGRPQQLNRVYGGEDSTDEMPWIVSIQ 63  
DB 7 LGGGCGTFTSLLLASTAALNAARIP-VPPACGRPQQLNRVYGGEDSTDEMPWIVSIQ 63

QY 69 KNGTHHCASLTSKRVYTAACFP-KDNINRKYLSVLGAMQOLNPGSRSKVYAWVE 127  
DB 69 KNGTHHCASLTSKRVYTAACFP-KDNINRKYLSVLGAMQOLNPGSRSKVYAWVE 127

DB 64 YGNNHCGGSLVSNMKNVSAACFPREHREAY-EVKIGAHQIDSYSDTYVHTVAQII 121  
DB 64 YGNNHCGGSLVSNMKNVSAACFPREHREAY-EVKIGAHQIDSYSDTYVHTVAQII 121

QY 128 PPHVYSWKGACADIALVLEKRSIOFSEVLPICLPDASIHLPMTWHWISGWSIQGY 187  
DB 128 PPHVYSWKGACADIALVLEKRSIOFSEVLPICLPDASIHLPMTWHWISGWSIQGY 187

DB 122 THSSYR-EEGSGGDIALLISPVYFSRIRICIPANASFPNGLHCTVGMGHVAFSV 180  
DB 122 THSSYR-EEGSGGDIALLISPVYFSRIRICIPANASFPNGLHCTVGMGHVAFSV 180

QY 188 PLPHPQTLQKLVPIIDSEVCSHLY-WRGAGGP--ITEDMLCAGYLGERDACLGDGSG 244  
DB 188 PLPHPQTLQKLVPIIDSEVCSHLY-WRGAGGP--ITEDMLCAGYLGERDACLGDGSG 244

DB 181 SLQTPRLPQLQLVPLISRETSCCLYNINAVPEPHTIQQDMLCAGYVKGKACQGDGSG 240  
DB 181 SLQTPRLPQLQLVPLISRETSCCLYNINAVPEPHTIQQDMLCAGYVKGKACQGDGSG 240

QY 245 PLMCQVDGKMLLAGITISWEGCAERNRPGVYSLSAHNSWVKYIQVQLR 295  
DB 245 PLMCQVDGKMLLAGITISWEGCAERNRPGVYSLSAHNSWVKYIQVQLR 295

DB 241 PLSCPEGIGIWLIVSWGDACGAPNRPGVYTLTSTYASWIIHHVAELDPR 291  
DB 241 PLSCPEGIGIWLIVSWGDACGAPNRPGVYTLTSTYASWIIHHVAELDPR 291

RESULT 5  
ID 091XC4 PRELIMINARY: PRT: 310 AA.

AC 091XC4: 01-DEC-2001 (TREMBlrel. 19, Created)  
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)  
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
DE SIMILAR TO DISTAL INTESTINAL SERINE PROTEASE.

OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;

SEQUENCE FROM N.A.  
Strausberg R.;  
Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL: BC010970; AAH10970.1; -;  
KW Protease.

SEQUENCE 310 AA; 33707 MW; FA126747DEAE0AB6 CRC64;

Query Match 33.2%; Score 571; DB 11; Length 310;  
Best Local Similarity 42.4%; Pred. No. 8.9e-45;  
Matches 117; Conservative 50; Mismatches 93; Indels 16; Gaps 9;

QY 21 LLLASTAILNAARIPV-PPACGRPQQLNRVYGGEDSTDEMPWIVSIQ-KNGTHHCAG 77  
DB 21 LLLASTAILNAARIPV-PPACGRPQQLNRVYGGEDSTDEMPWIVSIQ-KNGTHHCAG 77

QY 10 LLLL--QILFRAGDILPVCYGHSDKAGIKVGGQDALEQWQVSLWITEDG-HICGG 65  
DB 10 LLLL--QILFRAGDILPVCYGHSDKAGIKVGGQDALEQWQVSLWITEDG-HICGG 65

QY 78 SLTFRWYTAHCFKFNKLPYLSVLGAMQOLNPGSRSKVYAWVEPHTVSWKEG 137  
DB 78 SLTFRWYTAHCFKFNKLPYLSVLGAMQOLNPGSRSKVYAWVEPHTVSWKEG 137

QY 66 SLIEHWYLTAAHCFKFNKLPYLSVLGAMQOLNPGSRSKVYAWVEPHTVSWKEG 124  
DB 66 SLIEHWYLTAAHCFKFNKLPYLSVLGAMQOLNPGSRSKVYAWVEPHTVSWKEG 124

QY 138 ACADIALVLEKRSIOFSEVLPICLPDASIHLPMTWHWISGWSIQGYPLPHPQTLQK 197  
DB 138 ACADIALVLEKRSIOFSEVLPICLPDASIHLPMTWHWISGWSIQGYPLPHPQTLQK 197

QY 125 SSGDIATVQDPTLRPSPQ-FTPVCLPAAQPLPFGYCWMTGMCATGE--RDMAVSLGE 180  
DB 125 SSGDIATVQDPTLRPSPQ-FTPVCLPAAQPLPFGYCWMTGMCATGE--RDMAVSLGE 180

QY 198 LKVPILIDSEVCSHLY-WRG--AGQPTEDMLCAGYLGERDACLGDGSGPLMCQYDGA 253  
DB 198 LKVPILIDSEVCSHLY-WRG--AGQPTEDMLCAGYLGERDACLGDGSGPLMCQYDGA 253

QY 181 LAVPLIDSEDEKMYHVGSSLSGERIIQSDMLCAGYVEGQKSCQDGGPLVCSINSS 240  
DB 181 LAVPLIDSEDEKMYHVGSSLSGERIIQSDMLCAGYVEGQKSCQDGGPLVCSINSS 240

QY 254 WLAGIISWEGCAERNRPGVYSLSAHNSWVKYIQVQLR 289  
DB 254 WLAGIISWEGCAERNRPGVYSLSAHNSWVKYIQVQLR 289

QY 241 WTQVGTISWIGCARPYRPGVYTLTSTYASWIIHHVAELDPR 276  
DB 241 WTQVGTISWIGCARPYRPGVYTLTSTYASWIIHHVAELDPR 276

RESULT 6  
ID 096RZ8 PRELIMINARY: PRT: 321 AA.

AC 096RZ8: 01-DEC-2001 (TREMBlrel. 19, Created)  
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)  
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
DE HS TRANSMEMBRANE TRYPTASE, GENE NAME TMT, AF175522.1.

OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

OX NCBI\_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RA MEDLINE=21096910; PubMed=11157797;

RA Daxley R.J., Peden J.F., Lloyd C., Horsley S.W., Clark K.,  
RA Tufarelli C., Kearney L., Buckle V.J., Doggett N.A., Flint J.,  
RA Higgs D.R.;

RT "Sequence, structure and pathology of the fully annotated terminal 2  
RT Mb of the short arm of human chromosome 16."

RL Hum. Mol. Genet. 10:339-352(2001).

DR EMBL: AE006466; AAK61269.1; -;  
KW Transmembrane.  
SEQUENCE 321 AA; 33829 MW; FFF5089EDCAFC73D CRC64;

Query Match 33.1%; Score 570; DB 4; Length 321;  
 Best Local Similarity 41.1%; Pred. No. 1.2e-44;  
 Matches 123; Conservative 44; Mismatches 108; Indels 24; Gaps 8;

15 LGFTSLTLLASTALINARIPVPPACGKPOOLN---RVVGEDSTDEMPWIVSIQNG 71  
 3 LGACGILLTL---AVPGVSLRTLPGCGRPQVSDAGRIYGGHAAPAGAMPQASLRLRR 59  
 72 THHCAGSLTNRVITTAHCFKDNKRLYSVLGAMQLNGPSKQKGVAVWEHPV 131  
 60 VHVCGSLSPQWLTAAHCFSGSLNSSD-YOVHGLELITLSPHST---VROLIHSS 115  
 132 YSMKEGACADIALVRLERSIOFSEVRLPCLPDASIHLPNTHCWSISMGSIODGVPLPH 191  
 116 PSGGEGTSDIALVELSVPTLSSKRLPYCLPEASDDFCPGRCVMTGRTREEPRLP 175  
 192 POTOGLKVPITIDSEVCSHLVWKGAGOGPTEDMLCAGYLEGDERDACLSDSGPLMCOVD 251  
 176 PYSLEEVKYSVDTECRDYL-PGPGGSLTLPDMLCA---RQPGDACODSDSGPLVCOVN 231  
 252 GAMLGAGTISNMGCAERNRPVYISLSAHSRWEKIVQGVOLRRAGG---GALRAP 307  
 232 GAMVAGIVSMGCGGRPNRPVYTRVPYVNMIRRH1-----YASGSGSESGYPLRP 283

## RESULT 7

ID 090Y29 PRELIMINARY: PRT: 310 AA.

AC 090Y29; 01-MAY-2000 (TREMBlrel. 13, Created)  
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE DIGEST. INTESTINAL SERINE PROTEASE.  
 GN DISP OR DISP.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20246299; PubMed=10786627;  
 RA Shaw-Smith C.J., Coffey A.J., Leversha M., Freeman T.C., Bentley D.R.,  
 Walters J.R.;  
 RT "Characterization of a novel murine intestinal serine protease,  
 RT DISP.";  
 RT Biochim. Biophys. Acta 1490:131-136(2000).  
 -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1, ALSO KNOWN AS THE  
 TRYPsin FAMILY.  
 EMBL, AJ243866; CAB56465.1; -.  
 HSSP, P00763; IDPO.  
 DR MEROPS: S01.042; -.  
 DR MGD: MGI:133645; Disp.  
 DR InterPro: IPR001314; Chymotrypsin.  
 DR InterPro: IPR001254; Trypsin.  
 DR Pfam: PF00089; trypsin.1.  
 DR PRINTS: PR00722; CHYMOTRYPsin.  
 DR SMART: SM00020; TRYP\_SPC.1.  
 DR PROSITE: PS00240; TRYPsin\_DOM.1.  
 DR PROSITE: PS00134; TRYPsin\_HIS; UNKNOWN\_1.  
 DR PROSITE: PS00135; TRYPsin\_SER.1.  
 KW Hydrolyase; Protease; Serine protease.  
 SQ SEQUENCE 310 AA; 33701 MW; F828EC7F6D25303F CRC64;

Query Match 32.8%; Score 564; DB 11; Length 310;  
 Best Local Similarity 42.4%; Pred. No. 4e-44;  
 Matches 117; Conservative 48; Mismatches 95; Indels 16; Gaps 9;

21 LLLASTALINARIPV-PPACGKPOOLNRYVGEDSTDEMPWIVSIQNGTHHCAG 77  
 10 LLLT---QLTRAGDILPSVCGSHSDACKIVGGDADLEGQWPMQVOSIMTIEDG-HICGG 65

78 SLTNRVITTAHCFKDNKRLPYFVLLGAMQLNGPSKQKGVAVWEHPVYSKMG 137  
 66 SLIHEVAVLTAAHCFRSLN-PSFYHVKYVGLTLSTLEPHSTLVAVRNITFVHPTYLMDA 124  
 138 ACADIALVRLERSIOFSEVRLPCLPDASIHLPNTHCWSISMGSIODGVPLPHPTLCK 197  
 125 SSGDIALVOLDTPRLPSQ-FTPVCLPAAGRPPLTPGVTCWGTGAGTQ---RMAVYDQ 180  
 198 LKVPITIDSEVCSHLV-WRG---AGOGPTEDMLCAGYLEGDERDACLSDSGPLMCOVDA 253  
 181 LAVPLDSEDECKMYHTQSGSLSERITQSDMLCAGYVEBHIDSCGDSGPLYCSINS 240  
 254 WLAGTISNMGCAERNRPVYISLSAHSRWEKIV 289  
 241 WTQVITSMGICARPYRPGVYTRVPYVDMIORIL 276

## RESULT 8

ID 09XSM1 PRELIMINARY: PRT: 273 AA.

AC 09XSM1; 01-NOV-1999 (TREMBlrel. 12, Created)  
 DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE TRYPTASE (EC 3.4.21.59).  
 OS Ovis aries (Sheep).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 OC Bovidae; Caprinae; Ovis.  
 OX NCBI\_TaxID=9940;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX TISSUE=BONE MARROW;  
 RX MEDLINE=20308142; PubMed=10848900;  
 RA Scudamore C.L., McEuen A.R., Walls A.F., Miller H.R.P.;  
 RT "cDNA sequence of two sheep mast cell tryptases and the differential  
 RT expression of tryptase and sheep mast cell proteinase-1 in lung,  
 RT dermis and gastrointestinal tract.";  
 RL Clin. Exp. Allergy 30:818-863(2000).  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1, ALSO KNOWN AS THE  
 TRYPsin FAMILY.  
 CC EMBL: Y18223; CAB41986.1; -.  
 DR HSSP; P20231; IAAO.  
 DR MEROPS: S01.015; -.  
 DR InterPro: IPR001314; Chymotrypsin.  
 DR InterPro: IPR001254; Trypsin.  
 DR Pfam: PF00089; trypsin.1.  
 DR PRINTS: PR00722; CHYMOTRYPsin.  
 DR SMART: SM00020; TRYP\_SPC.1.  
 DR PROSITE: PS00240; TRYPsin\_DOM.1.  
 DR PROSITE: PS00134; TRYPsin\_HIS; UNKNOWN\_1.  
 DR PROSITE: PS00135; TRYPsin\_SER.1.  
 KW Hydrolyase; Serine protease. TRYPTASE.  
 FT CHAIN 29 273  
 SQ SEQUENCE 273 AA; 30156 MW; 88A19DC779053898 CRC64;

Query Match 32.5%; Score 559.5; DB 6; Length 273;  
 Best Local Similarity 40.9%; Pred. No. 8.8e-44;  
 Matches 113; Conservative 53; Mismatches 95; Indels 15; Gaps 7;

21 LLLASTALINARIPVPPACGKPOOLNRYVGEDSTDEMPWIVSIQ---KNGTHHCAG 77  
 2 LHLTALLLTLSTL-VSAPGPGQALQSGILGKEAPGSRPMQVSLRVADQYWRHCCG 59  
 78 SLTNRVITTAHCFKDNKRLPYFVLLGAMQLNGPSKQKGVAVWEHPVYSKMG 137  
 60 SLIHPQWLTAAHCFGLQELQPSDFRVOLREOHL---YQDRLLPISRVIPRPHYVWEN 116  
 138 ACADIALVRLERSIOFSEVRLPCLPDASIHLPNTHCWSISMGSIODGVPLPHPTLCK 197  
 117 G-ADIALQLEEPVYSICHVAPVTLPPASETFPPGSCQWVTKGMDNGRPLPPYPLKO 175

Query Match	31.2%;	Score 537;	DB 11;	Length 799;
-------------	--------	------------	--------	-------------

	Best Local Similarity	40.7%;	Pred. No. 4,1e-41;	
	Matches 103;	Conservative 54;	Mismatches 82;	Indels 14; Gaps 6;
QY	41 CGKQQLNRVVGGEHDSIDSEWPIVTISQKNGTHHCAGSLTTSRWITAAHCF-KDNLNKP	99		
Dd	556 CGLGGLSSRIIVGVGSSEGEWPQASIQIIGRHICGALLADRWITAAHCFQEDSMASP	615		
QY	100 YLFVSLLGAQMOLN--PGSRQKVGYAWVEHPHYVMKMGACADIALVLERSIQFSERY	157		
Dd	616 KLMVFELGKKRQNSRNRPGEYSFKVSKLELHP--YHEEDSHDYDALQLDLDPHYSATV	672		
QY	158 LPILCPDASTHLPPNTHCWTISGWGSIODGVPLPHPTLOKLKVPPIIDSEVCSHLYWRGAG	217		
Dd	673 RPYCLPARSHFFEPGGCHCWITGWGAREGPVSN--TLQKYDVQLVLPQDLCEAT-----	725		
QY	218 QGFPTEDMTCAGYLEEERACLGDSGGPLMC-QVDCAWLIAITISKBGCARNPFGYVI	276		
Dd	726 RLSVSPMLLCAGRKRRKDACQDGSGGLVCRBPSGRMFLAGLVSWGLCGRPNPFYYT	785		
QY	277 TLSAHRSWEKVI 289			
Dd	786 RVTRVINMIQQVL 798			
RESULT	10			
ID	Q29464	PRELIMINARY:	PRT:	237 AA.
AC	Q29464;			
DT	01-NOV-1996 (TREMBLREL. 01, Created)			
DT	01-NOV-1996 (TREMBLREL. 01, Last sequence update)			
DE	01-DEC-2001 (TREMBLREL. 19, last annotation update)			
DE	TRYPLASE (EC 3.4.21.59) (FRAGMENT).			
OS	Bos taurus (Bovine).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea;			
OC	Bovidae; Bovinae; Bos.			
OX	NCBL_TaxId=9913;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
KC	TISSUE=LIVER CAPSULE;			
RX	MEDLINE=96203914; PubMed=8620861;			
RA	Pallao M., Gambacorta A., Fiorucci L., Mignogna G., Barra D.,			
RA	Ascoli F.;			
RT	"cDNA cloning and primary structure of trypsinase from bovine mast cells			
RT	and evidence of the expression of bovine pancreatic trypsin inhibitor			
RT	mRNA in the same cells.";			
RL	Eur. J. Biochem. 237:100-105(1996).			
CC	-!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE			
CC	TRYPSIN FAMILY			
DR	EMBL; X94982; CA64438.1; -.			
DR	HSSP; P20231; IAAO.			
DR	MEROPE; S01.143; -.			
DR	InterPro: IPRO01314; Chymotrypsin.			
DR	InterPro: IPRO01254; Trypsin.			
DR	Pfam: PF00089; trypsin_1			
DR	PRINTS; PR00722; CHYMOTRYPSIN.			
DR	SMART; SM00020; TRYP_SPC; 1.			
DR	PROSITE; PS50240; TRYPSIN_DOM; 1.			
DR	PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.			
KW	Hydrolase; serine protease.			
FT	NON_TER 1			
SO	SEQUENCE 237 AA; 26550 MW; EA537A75294EFFBA CRC64;			
Query Match	31.2%;	Score 536;	DB 6;	Length 237;
Best Local Similarity	42.6%;	Pred. No. 1.1e-41;		
Matches 103;	Conservative 48;	Mismatches 75;	Indels 16;	Gaps 7;
QY	59 SEWMIVSI---QKNGTNHCAGSLTTSRWITTAHCFKDNLNKPYLFVSLLGAQMOLNPG	115		
Dd	2 SQPMQVSLAKSVYMRHRCGGSLIHQVWLVAHAHVGVGEHPGPSFYVOLREOHU---Y	58		
XY	116 SRQKVVANAVEBHP-VYSWKEGACADIALVLRERSIQFSERVLPCLDPASIIHPNTH	174		

```

Db      : | : : | | | | | | | | | | : : : | | : | | |
59 YODLLPISRIIPHPNCSYKNG--ADIALLELDKLVNISMHYOPVTLPPSESEFPFGTO 116

QY      175 CWISGWSIODGVPLPHPOTILOKLVPIIDSEVCSHLYMRGAGGP----ITEDMLCAGY 230
      | : | | : : | | | | : : | | : : | | : | | | |
Db      117 CWTGNGNVDNGRRLPPFPPLKQYKVPVENSVCGDRKYHSGLSTGDNVPIVREDMLCAG- 175
      | : | | : : | | | | : : | | : : | | : | | | |
QY      231 LEGERDACLIGSGGPIKCOVDGAWLAGIISWEGCAERNRPGVYISLSAHSRWEKIV- 289
      | : | | : : | | | | : : | | : : | | : | | | |
Db      176 -DSGRNFCQGDGGGPLVCKVNGTWTLOAGVYSWGDCAKPNRPGIYTRVTSYLDWIHQYVP 234

QY      290 QG 291
      | |
Db      235 QG 236

```

Search completed: August 13, 2002, 08:51:08  
 time: 226 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 13, 2002, 08:45:12 ; Search time 13.05 Seconds  
(without alignments)  
593.327 Million cell updates/sec

Title: US-10-040-803-7

Perfect score: 1720  
Sequence: 1 MVSQAPALGGGCGTFTS.....AOGGALFAPSQSGAARS 317

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 08  
Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents, AA:\*

1: /cgn2\_6/ptodata/2/1aa/5A.COMB.pep:\*  
2: /cgn2\_6/ptodata/2/1aa/5B.COMB.pep:\*  
3: /cgn2\_6/ptodata/2/1aa/6A.COMB.pep:\*  
4: /cgn2\_6/ptodata/2/1aa/6B.COMB.pep:\*  
5: /cgn2\_6/ptodata/2/1aa/PCRTUS.COMB.pep:\*  
6: /cgn2\_6/ptodata/2/1aa/Backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	581.5	33.8	299	4	US-08-944-483-66
2	554	32.2	276	2	US-09-016-366A-15
3	554	32.2	276	2	US-08-978-404B-21
4	538.5	31.3	274	4	US-08-978-404B-5
5	531.5	30.9	314	4	US-09-008-271A-3
6	530.5	30.8	274	2	US-08-916-366A-21
7	530.5	30.8	274	2	US-08-978-404B-16
8	530.5	30.8	275	2	US-09-016-366A-17
9	530.5	30.8	275	2	US-08-978-404B-12
10	529.5	30.8	273	2	US-09-016-366A-19
11	529.5	30.8	273	2	US-08-978-404B-14
12	519.5	30.2	273	2	US-08-978-404B-8
13	519.5	30.2	273	2	US-08-978-404B-6
14	518	30.1	273	2	US-08-978-404B-3
15	514.5	29.9	267	2	US-09-016-366A-23
16	514.5	29.9	267	2	US-08-978-404B-18
17	511.5	29.7	245	4	US-08-944-483-69
18	511.5	29.7	245	4	US-08-944-483-65
19	508	29.5	304	4	US-09-079-970A-5
20	506.5	29.4	245	4	US-09-079-970A-6
21	473.5	27.5	248	4	US-08-944-483-63
22	470.5	27.4	238	4	US-08-944-483-64
23	469.5	27.3	638	2	US-08-681-151-3
24	468.5	27.2	435	4	US-09-008-271A-6
25	466	27.1	454	4	US-09-518-046-2
26	463	26.9	416	2	US-09-000-846-2
27	462	26.9	492	4	US-09-342-749-2

28	453	26.3	283	3	US-08-807-151-1	Sequence 1, Appli
29	453	26.3	283	4	US-09-478-957-1	Sequence 1, Appli
30	449	26.1	791	1	US-08-643-219-1	Sequence 1, Appli
31	449	26.1	791	3	US-08-851-350-1	Sequence 1, Appli
32	447	26.0	814	1	US-08-750-711-1	Sequence 1, Appli
33	446	25.9	546	6	5200340-6	Patent No. 5200340
34	446	25.9	790	1	US-08-469-658-54	Sequence 54, Appl
35	446	25.9	790	2	US-08-469-658-54	Sequence 54, Appl
36	446	25.9	791	2	US-09-131-995-1	Sequence 1, Appli
37	446	25.9	791	2	US-08-832-087B-1	Sequence 1, Appli
38	446	25.9	791	4	US-09-132-154-1	Sequence 1, Appli
39	446	25.9	810	1	US-07-854-603-2	Sequence 2, Appli
40	446	25.9	810	1	US-08-147-000B-29	Sequence 29, Appli
41	446	25.9	810	4	US-09-086-514-1	Sequence 1, Appli
42	446	25.9	810	6	5200340-8	Patent No. 5200340
43	445.5	25.9	256	2	US-09-027-337-3	Sequence 3, Appli
44	443.5	25.8	798	1	US-08-200-900A-2	Sequence 2, Appli
45	443.5	25.8	798	5	PCT-US94-00616-2	Sequence 2, Appli

## ALIGNMENTS

RESULT 1  
US-08-944-483-66  
; Sequence 66, Application US/08944483  
; Patent No. 6232456  
; GENERAL INFORMATION:  
; APPLICANT: COHEN, MAURICE  
; APPLICANT: COLPITTS, TRACEY L.  
; APPLICANT: FRIEDMAN, PAULA N.  
; APPLICANT: GRANADOS, EDWARD N.  
; APPLICANT: KLASS, MICHAEL R.  
; APPLICANT: RUSSELL, JOHN C.  
; APPLICANT: STEWART, KENT D.  
; APPLICANT: STROUPE, STEVEN D.  
; TITLE OF INVENTION: NOVEL SERINE PROTEASE REAGENTS  
; TITLE OF INVENTION: AND METHODS USEFUL FOR DETECTING AND TREATING DISEASES  
; NUMBER OF SEQUENCES: 76  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: Abbott Laboratories  
; STREET: 100 Abbott Park Road  
; CITY: Abbott Park  
; STATE: IL  
; COUNTRY: USA  
; ZIP: 60064-3500  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FASTSEQ for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/944,483  
; FILING DATE:  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Becker, Cheryl L.  
; REGISTRATION NUMBER: 35,441  
; REFERENCE/DOCKET NUMBER: 6183-US 01  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 847/935-1729  
; TELEFAX: 847/938-2623  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 66:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 299 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear

MOLECULE TYPE: No. 62324566  
US-08-944-483-66

Query Match 33.8%; Score 581.5; DB 4; Length 299;  
Best Local Similarity 45.4%; Pred. No. 6.5e-50;  
Matches 113; Conservative 42; Mismatches 89; Indels 5; Gaps 4;

QY 50 VVGEDSDTSEMPWIVSIQKNGTHHCAGSLTISRMTAAHCFKDNLNKPYLSVLLGAW 109  
DB 1 ITGSSANAGMOWOVITTYEGVAVCGSLVSEQWVLSAHCPESEHKE-AVEVKLGAH 59  
QY 110 QINGRSRQKVGAVWEPHVPYVSKBACADIALVRLERSIQSESEKVLPICLPDASIH 169  
DB 60 QLDVSEDAKVTSLKDLIPHSY-LOESQSGDIALDLISREITPSRYIRPICLPANASF 118  
QY 170 PPNHWCWISGWSIODGVPILPHPQTLQKVPILIDSEVCSHLWYRGA-GQGP--ITEDML 226  
DB 119 PNGHCTVYTGWHAAPSLSLTPKPLDOLLEVPILSRETGCLNYMDAKPREPHFVQEDMV 178  
DB 227 CAGYLEGERDACLGDGSGPLMCQVYGAWLLAGIISWEGCAERNRPVYISLSAHSRWE 286  
DB 179 CAGYEGSKDAGCGSGPLSCPYEGWLYTGIYSWGDACGARNRPVYITLASSYASWIQ 238  
QY 287 KIVGCVQLR 295  
DB 239 SKVTELQPR 247

RESULT 2  
US-09-016-366A-15  
Sequence 15, Application US/09016366A  
Patent No. 5955431

GENERAL INFORMATION:  
APPLICANT: Stevens, Richard L.  
APPLICANT: Huang, Chifu  
TITLE OF INVENTION: MAST CELL PROTEASE PEPTIDE  
TITLE OF INVENTION: INHIBITORS  
NUMBER OF SEQUENCES: 65  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Wolf, Greenfield & Sacks, P.C.  
STREET: 600 Atlantic Avenue  
CITY: Boston  
STATE: MA

COUNTRY: U.S.A.  
ZIP: 02210-2211  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS

SOFTWARE: FASTSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: 60/037,090  
FILING DATE: January 30, 1998

CLASSIFICATION: 530  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: 60/037,090  
FILING DATE: 05-FEB-1997

ATTORNEY/AGENT INFORMATION:  
NAME: Plumer, Elizabeth R.  
REGISTRATION NUMBER: 36,637  
REFERENCE/DOCKET NUMBER: B0801/7093  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-720-3500  
TELEFAX: 617-720-2441

TELEX:  
INFORMATION FOR SEQ ID NO: 15:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 276 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein

US-09-016-366A-15

Query Match 32.2%; Score 554; DB 2; Length 276;  
Best Local Similarity 41.6%; Pred. No. 3.2e-47;  
Matches 116; Conservative 48; Mismatches 93; Indels 22; Gaps 9;

QY 21 LILLASTALINARIPVP-PACGKPDQNLNRYVVGSDSDSEMPWIVSIQ--KNGTHICA 76  
DB 6 LILLASTALINARIPVP-PACGKPDQNLNRYVVGSDSDSEMPWIVSIQ--KNGTHICA 76  
QY 77 GSLTISRMTAAHCFKDNLNKPYLSVLLGAWQ--GNPGRSQKVGAVWEPHVPYVSW 134  
DB 62 GSLTISRMTAAHCFKDNLNKPYLSVLLGAWQ--GNPGRSQKVGAVWEPHVPYVSW 134  
QY 135 KEACADIALVRLERSIQSESEKVLPICLPDASIHLPNNHWCWISGWSIODGVPILPHPQ 194  
DB 117 AEQG-ADVALLLEVPVNSTHHPISLTPASTFPFGISCVWGTGDIQNDNEPLPPYV 175  
QY 195 LQKVPYVNSLDCRKYHTGLYTGDPFIVHDMCLAG--NTRRDSGCGSGGPLVCKV 233  
DB 176 LQKVPYVNSLDCRKYHTGLYTGDPFIVHDMCLAG--NTRRDSGCGSGGPLVCKV 233  
QY 251 DGAWLLAGIISWEGCAERNRPVYISLSAHSRWEKIV 289  
DB 234 KGTWLAGVSWEGCAQPNKPGIYTRVYITLDMIRIV 272

RESULT 3  
US-08-978-404B-21  
Sequence 21, Application US/08978404B  
Patent No. 5968782

GENERAL INFORMATION:  
APPLICANT: Stevens, Richard L.  
APPLICANT: Huang, Chifu  
TITLE OF INVENTION: MAST CELL PROTEASE THAT CLEAVES  
TITLE OF INVENTION: FIBRINOGEN  
NUMBER OF SEQUENCES: 74  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Wolf, Greenfield & Sacks, P.C.  
STREET: 600 Atlantic Avenue  
CITY: Boston  
STATE: MA

COUNTRY: U.S.A.  
ZIP: 02210-2211  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS

SOFTWARE: FASTSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/978,404B  
FILING DATE: 25-NOV-97

CLASSIFICATION: 435  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: 60/032,354  
FILING DATE: 04-DEC-1996

ATTORNEY/AGENT INFORMATION:  
NAME: Plumer, Elizabeth R.  
REGISTRATION NUMBER: 36,637  
REFERENCE/DOCKET NUMBER: B0801/7090  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-720-3500  
TELEFAX: 617-720-2441

TELEX:  
INFORMATION FOR SEQ ID NO: 21:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 276 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: NO. 5968782e  
US-08-978-404B-21



TYPE: amino acid

; MOLECULE TYPE: NO. 5968782e

MOLECULE TYPE: NO. 3900







Tue Aug 13 16:27:46 2002

us-10-040-803-7.rai

---

